

مُقَدِّمةٌ في حَوْسَبَة اللُّغَـــة العَرَبِيَّة

مباحث لغوية 00

تحرير

د. مُحســـن رَشـــوان   د. المُعتزّ بالله السَّعيد

الباحثون:

د. مُحســـن رَشـــوان   د. أحمـــــد راغِـــب

د.مُحمَّـــد عَطيَّـــــة   د. سامِـــح الأنصـــاري

د. المُعتزّ بالله السَّعيد

مباحث لغوية ٥٥

مُقَدِّمَةٌ في
حَوْسَبَة اللُّغَـــة العَرَبِيَّة

تحرير

د. مُحسـن رَشـوان   د. الـمُعتزّ بالله السَّعيـد

الباحثون:

د. مُحسـن رَشـوان   د. أحمـد راغِب

د. مـُحمَّـد عَطيَّـة   د. سامِـح الأنصـاري

د. الـمُعتزّ بالله السَّعيـد

١٤٤١هـ - ٢٠١٩م



مركز الملك عبدالله بن عبدالعزيز الدولي
لخدمة اللغة العربية
King Abdullah Bin Abdulaziz Int'l Center for
The Arabic Language

مُقَدِّمةٌ في حَوْسَبَة اللُّغَة العَرَبِيَّة

الطبعة الأولى
١٤٤١ هـ - ٢٠١٩ م

المملكة العربية السعودية - الرياض
ص.ب ١٢٥٠٠ الرياض ١١٤٧٣
هاتف:٠٠٩٦٦١١٢٥٨٧٢٦٨ - ٠٠٩٦٦١١٢٥٨١٠٨٢
البريد الإليكتروني: nashr@kaica.org.sa

مركز الملك عبدالله بن عبدالعزيز الدولي لخدمة اللغة العربية، ١٤٤١هـ.
فهرسة مكتبة الملك فهد الوطنية أثناء النشر
رشوان، محسن
مقدمة في حوسبة اللغة العربية./ محسن رشوان؛ المعتز بالله السعيد .-الرياض، ١٤٤٠هـ
..ص؛ .. سم
ردمك: ٨-٥١-٨٢٢١-٦٠٣-٩٧٨
١ - اللغة العربية - معالجة البيانات أ. السعيد ، المعتز بالله (مؤلف مشارك) ب. العنوان
ديوي ٤١٠,٢٨٥   ١٤٤٠/١٠١٦٥
رقم الإيداع: ١٤٤٠/١٠١٦٥
ردمك: ٨-٥١-٨٢٢١-٦٠٣-٩٧٨

التصميم والإخراج
وجوه WOJOOOH
دار وجوه للنشر والتوزيع
Wojooh Publishing & Distribution House
www.wojoooh.com
المملكة العربية السعودية - الرياض
الهاتف:4562410 الفاكس:4561675
للتواصل والنشر:
info@wojoooh.com


بسم الله الرحمن الرحيم

فهرس الكتاب

كلمة المركز

يعمل المركز في مجال البحث العلمي ونشر الكتب مستهدفاً التركيز على المجالات البحثية التي ما زالت بحاجة إلى تسليط الضوء عليها، وتكثيف البحث فيها، ولفت أنظار الباحثين والجهات الأكاديمية إلى أهمية استثمارها بمختلف وجوه الاستثمار، وذلك مثل مجال (التخطيط اللغوي) و (العربية في العالم) و(الأدلة والمعلومات) و (تعليم العربية لأبنائها أو لغير الناطقين بها) إلى غير ذلك من المجالات، وإن من أهم مجالات البحث المستقبلية في اللغة العربية مجال (العربية والحوسبة ، والذكاء الاصطناعي) حيث إن حياة اللغات ومستقبلها مرهونة بمدى تجاوبها مع التطورات التقنية والعالم الافتراضي، وكثافة المحتوى الالكتروني المكتوب، وهو ما يشكّل تحديا حقيقيا أمام اللغات غير المنتجة للمعرفة أو للتقنية.

وقد عمل المركز على تسليط الضوء على هذا المجال التخصصي؛ مستعينا بالكفاءات القادرة من المهتمين بالتخصص البيني (بين اللغة والحاسوب) مقدّرا جهودهم، وهادفاً إلى نشرها، وتعميم مبادئها، راغباً أن يكون هذا المسار العلمي مقررا في الجامعات في كلية العربية والحاسوب، ومجالا بحثيا يقصده الباحثون الأكاديميون، والجهات البحثية العربية.

وقد أصدر المركز سابقا ستة عشر كتاباً مختصا في (حوسبة العربية) وفي الإفادة من (المدونات اللغوية) في الأبحاث العربية، ويحتفل بإصدار سبعة كتب جديدة مختصة في (حوسبة العربية والذكاء الاصطناعي)، ويقدمها للقارئ العربي، وللجهات الأكاديمية؛ للإفادة منها في مناهج التعليم والبناء عليه، وهذه الكتب السبعة هي: (العربيَّة والذَّكاء الاصطناعيّ، تطبيقات الذكاء الاصطناعي في خدمة اللغة العربية، خوارزميات الذكاء الاصطناعي في تحليل النص العربي، مقدمة في حوسبة اللغة العربية، الموارد اللغوية الحاسوبية، المعالجة الآلية للنصوص العربية، تطبيقات أساسية في المعالجة الآلية للغة العربية).

ويشكر المركز السادة مؤلفي الكتب، ومحرريها، لما تفضلوا به من عمل علمي رصين، وأدعو الباحثين والمؤلفين إلى التواصل مع المركز لاستكمال المسيرة، وتفتيق فضاءات المعرفة.

وفق الله الجهود وسدد الرؤى.

الأمين العام

أ. د. محمود إسماعيل صالح

مقدمة

مع الطَّفرة المعلوماتيَّة الَّتي شَهِدَها القَرنُ الحادي والعشرون، أمكنَ إخضاعُ الآلة لمُعالَجة نُصُوص اللُّغات الطَّبيعيَّة. وساعدَ ذلكَ في توفير الوقت والجُهد البشريّ المبذول؛ خُصُوصًا حينَ يتعلَّقُ الأمرُ بمجموعاتٍ كبيرةٍ من نُصُوص اللُّغة.

وقد حَظِيَت اللُّغةُ العربيَّة بعناية الباحثينَ في مُعالَجة اللُّغات الطَّبيعيَّة في المشرق والمغرب؛ وساعدَت بعضُ خصائص اللُّغة العربيَّة على تطوير أدواتٍ ناجعةٍ لمُعالَجتِها آليًّا عبرَ عدَّة مُستوياتٍ، ومن هذه الخصائص الطَّبيعةُ القياسيَّة للصَّرف العربيّ الَّتي مَكَّنَت من تطوير أدوات التَّحليل الصَّرفيّ. وفي الوقت ذاته، مثَّلَت بعضُ خصائص العربيَّة تحدياتٍ أمامَ حوسَبَتِها، مثل طبيعة نظامها الكتابيّ، وطبيعة بنيتها التَّركيبيَّة.

وسعيًا إلى الوقوف على طرائق المُعالَجة الآليَّة للُّغة العربيَّة والتَّحدِّيات الَّتي تُواجهُ حَوسَبَتَها، فإنَّنا نُقدِّمُ للقارئ العربيّ كتاب (مُقدِّمة في حَوسَبة اللُّغة العربيَّة)، آملينَ أن يُسهمَ في إثراء المكتبة العربيَّة الَّتي تُعاني نقصًا كبيرًا في مصادر مُعالَجة اللُّغة العربيَّة وتقنياتها. ولعلَّنا بذلكَ نفتحُ البابَ أمامَ الباحثينَ لطرق هذا الميدان وتنمية أساليب البحث فيه.

ولما كانَ الهدفُ من هذا الكتاب توجيهَ القارئ العربيّ إلى مبادئ حوسَبة اللُّغة ومُقدِّماتِها، رأينا أن ننطلقَ من مُستويات التَّحليل اللُّغويّ الأساسيَّة المُتعارَف عليها بينَ الباحثينَ في عُلُوم اللُّغة؛ ونعني المُستويات الخمسة (الصَّوتيّ، والصَّرفيّ، والتَّركيبيّ، والدِّلاليّ، والمُعجميّ) مع التَّقديم لذلكَ بمدخلٍ إلى حَوسَبة اللُّغة. لهذا، فقد قسَّمنا الكتابَ إلى ستَّة فُصُولٍ، على النَّحو الآتي:

- الفصل الأوَّل: مدخل إلى حَوسَبة اللُّغة؛ يُعنى بمفهوم حَوسَبة اللُّغة وأبرز الاصطلاحات المُستخدمة في التَّعبير عن هذا المَيدان؛ ثُمَّ يعرضُ للتَّحدِّيات الَّتي تُواجهُ مُعالَجَةَ اللُّغات الإنسانيَّة، واللُّغة العربيَّة على وجه الخُصُوص. ويُقدِّمُ الفصلُ تمهيدًا حولَ تقنيات اللُّغات الإنسانيَّة ووظائفِها، ويعرضُ كذلكَ تعريفًا ببعض المُؤسَّسات المَعنيَّة بحوسَبَة اللُّغة العربيَّة.
- الفصل الثاني: التَّحليل الصَّوتيّ؛ ويَعرِضُ دراسةً تحليليَّةً للصَّوت اللُّغويّ في العَصر الحديث، مُقدِّمًا لذلكَ بأهمِّ المُصطلحات المُستَخدَمة في المُعالَجَة الآليَّة للصَّوت اللُّغويّ ومفاهيمها. ويعرضُ هذا الفصلُ لكيفيَّة التَّحليل الصَّوتيّ الحاسوبيّ، من خلال الإبانة عن الأجهزة الأساسيَّة لحوسَبة الصَّوت اللُّغويّ، وكذلك الأدوات والبرمجيَّات المُساعدة في تحليل أصوات اللُّغة.
- الفصل الثَّالث: التَّحليل الصَّرفيّ الآليّ للمُفردات العربيَّة؛ يُعنى هذا الفصلُ بطبيعة البناء الصَّرفيّ في اللُّغة العربيَّة، ويُقَدِّمُ تصوُّرًا لحجم حصيلة المُفردات العربيَّة. ويُقدِّمُ تعريفًا للتَّحليل الصَّرفيّ الآليّ وتطبيقاته، والالتباس الصَّرفيّ وأساليب إزالته. ويعرضُ كذلكَ للموارد اللُّغويَّة اللَّازمة لإزالة الالتباس الصَّرفيّ.
- الفصل الرَّابع: التَّحليل التَّركيبيّ؛ يُقَدِّمُ هذا الفصلُ تمهيدًا حولَ علاقة منظومة النَّحو بمنظومَتَي الصَّرف والمُعجم، ثُمَّ مدخلًا إلى تركيب الجُملة في اللُّغة، والمُعالَجَة الآليَّة لها. ويعرضُ كذلكَ لأساليب بناء المُحلِّلات النَّحويَّة [التَّركيبيَّة] واتِّجاهاتها ودقَّتِها، مع التَّمثيل عليها. ويعرضُ الفصلُ أيضًا لبعض النِّقاط البحثيَّة الَّتي تُساعدُ في تطوير مُعالَجَة التَّركيب في اللُّغة العربيَّة.

- الفصل الخامس: التَّحليل الدِّلاليّ؛ يُقدِّمُ مدخلًا للتَّحليل الدِّلاليّ في اللُّغات الطَّبيعيَّة عمومًا، وفي اللُّغة العربيَّة على وجه الخُصُوص. ويعرضُ لشبكات الدِّلالات المُعجميَّة، كما يعرضُ للالتباس الدِّلاليّ وأساليب إزالتِه، ويعرضُ كذلكَ لتطبيقات التَّحليل الدِّلاليّ والعَنونة الدِّلاليَّة للمُدوَّنات اللُّغويَّة.
- الفصل السَّادس: التَّحليل المُعجميّ الآليّ؛ ويأتي هذا الفصل تتمَّةً لُمستويات التَّحليل اللُّغويّ؛ حيثُ يُقدِّمُ لمفهوم التَّحليل المُعجميّ، ويعرضُ لمُكوِّنات المُعجم اللُّغويّ، سواءٌ أكانت مُكوِّناتٍ أساسيَّة أم ثانويَّة. ويُقدِّمُ الفصلُ مدخلًا للمُعالَجة المُعجميَّة الآليَّة في مراحل الصِّناعة المُعجميَّة، جمعًا وتحريرًا ونشرًا؛ ويعرضُ كذلك لأدوات المُعالَجة الآليَّة لمُكوِّنات المُعجم اللُّغَويّ، مع التَّركيز على المُعجم العربيّ.

إنَّنا نأمل أن يجدَ القارئُ الكريمُ في هذا الكتاب مُقدِّمةً وافيةً لحوسَبة اللُّغة العربيَّة، تُعينُهُ على فهم منطق الآلة في المُعالَجة الآليَّة لمُفرَدات اللُّغة وتراكيبها، بصورَتَيها: المنطوقة والمكتوبة. ونرجو كذلكَ أن يكونَ هذا الكتابُ مفتاحًا للباحثينَ في ميادين اللُّغة العربيَّة للتَّفكير في تطوير أساليب مُعالَجتِها وتقعيدِها وتيسير تعليمها وتعلُّمها.

نسألُ الله تعالى أن يتقبَّلَ هذا الجهدَ بالذِّكر الحَسَنِ والأجرِ الجزيل، وأن يجعلَه من العلمِ الَّذي ينفعُ أصحابَه بعد مماتهم.

ربَّنا عليكَ توكَّلنا وإليكَ أنَبنا وإليكَ المصِير.

الـمُحرِّران

الفصل الأول
مدخلٌ إلى حوسبة اللُّغة

د. مُحسِن رَشوان

١- مُستويات التَّحليل اللُّغَوِيّ.
٢- التَّحدِّيات الَّتي تواجه تقنيات اللُّغات الإنسانِيَّة.
٣- التَّحدِّيات الَّتي تُواجه معالجة اللغة العربية.
٤- التَّعريف بتقنيات اللُّغات الإنسانيَّة.
٥- من المؤسَّسات المعنِيَّة بحوسبة اللُّغة العربيَّة.
٦- أفكارٌ بحثيَّة لأطرُوحاتٍ علميَّةٍ مُستَقبليَّة.

تمهيد

يُعنى هذا الكتاب بحوسبة اللُّغة العربيَّة وما يتَّصِلُ بها من مُقَدِّمات. وحوسبة اللُّغة علمٌ حديثٌ نسبِيًّا، ظهرَ في مطلع النِّصف الثَّاني من القرن العشرين باعتبارِه علمًا بينِيًّا، يُعتَمَدُ فيه على الحاسوب لدراسة اللُّغات الإنسانيَّة وفهم طبيعتها.

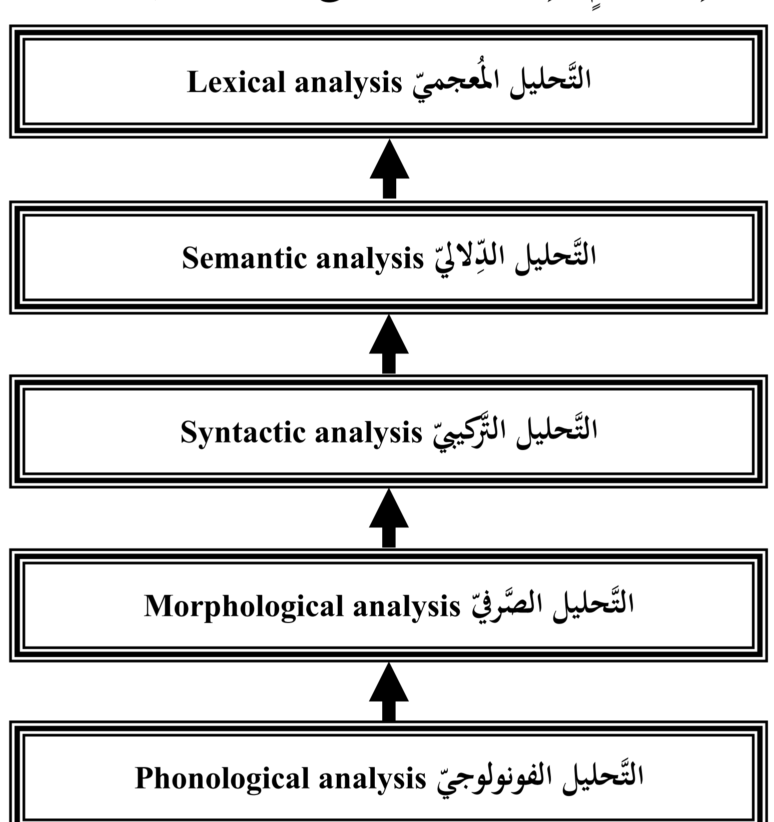
ويُشارُ إلى هذا العلم بمُسَمَّياتٍ أخرى، لعلَّ أهمَّها: اللِّسانيَّات الحاسُوبيَّة (Computational Linguistics - CL)، ومُعالَجة اللُّغات الطَّبيعِيَّة (Natural Language Processing - NLP)، وتقنيات اللُّغة الطَّبيعِيَّة (الإنسانِيَّة) (Human Language Technologies - HLT). ومع تعدُّد هذه المسَمَّيات وتفاوُت أساليب دراستها بين اللُّغَوِيِّين والحاسوبيِّين، إلاَّ أنَّها تدورُ جميعًا في دائرةٍ واحدة، هي دائرةُ ذلك العلم الَّذي تُوَجَّهُ من خلاله أنظمة الحاسوب إلى فهم لُغة الإنسان ومُحاكاة الذَّكاء البَشَرِيّ.

وتقومُ حوسبةُ اللُّغة على ثلاثة محاورَ أساسيَّة، هي:

- تقنيات معالجة النُّصوص (Text Processing). ومن أمثلتِها: التَّرجمة الآليَّة، والتَّلخيص الآليّ، والتَّنقيب في النُّصوص.
- تقنيات معالجة الكلام المنطوق (Speech Processing). ومن أمثلتِها: التَّعَرُّف الآليّ على الكلام المنطوق، وتحويل النَّصّ المكتوب إلى كلام منطوق.
- تقنيات مُعالَجة الصُّوَر (Image Processing). ومن أمثلتِها: التَّعَرُّف الآليّ على الكتابة (Optical Character Recognition - OCR).

١ - مُستويات التَّحليل اللُّغَوِيّ

نودُّ أن نُشِيرَ - بدايةً - إلى ما جرى عليه الباحثون في حوسبة اللُّغة من ترتيب مراحل التَّحليل اللُّغَوِيّ في سُلَّمٍ لُغَوِيٍّ على النَّحو الموَضَّح في الشَّكل التَّالي:



الشَّكل ١-١: مُستويات التَّحليل اللُّغَوِيّ وفقَ ما جرى عليه الباحثون.

ووفقًا للعديد من اللُّغات الإنسانيَّة، يجدُرُ بنا أن نُشيرَ إلى وُجود تداخُلٍ وارتباطٍ بين مُستويات التَّحليل اللُّغَوِيّ؛ وإذا اعتبَرنا أنَّ هذه المستويات تُمَثِّلُ مجموعةً من الدَّرَجات المتعاقِبة، فباستطاعتنا أن نَتَبَيَّنَ طبيعَتَها وماهِيَّتَها على النَّحو التَّالي:

١ , ١ – التَّحليل الفونولوجيّ (Phonological analysis)

ويتمُّ في هذه الدَّرجة حسم طريقة نُطق الكلمة مع مُراعاة الحُرُوف الَّتي لا تُنطَق أو الَّتي تُنطَق على غير أصلِها (كالحُرُوف المدغَمة والحُرُوف الَّتي يلتقي فيها السَّاكنان). وتُعَدُّ علومُ التَّجويد الوسيلةَ الأساسيَّة الَّتي تُساعد على معرفة قواعد نُطق القرآن الكريم، مع ملاحظة التَّبايُن بينها وبينَ قواعد نُطق العربيَّة الفُصحى في عُمُومِها.

وتُواجهُنا في العربيَّة – خُصُوصًا – صعوبةٌ تكمنُ في عدم استخدام الحركات القصيرة (short vowels)، وهذا يجعل إنجازَ آليَّة النُّطق الآليَّ لكلمات العربيَّة مُهِمَّةً صعبة مُقارَنةً بنظائِرِها من اللُّغات الَّتي تُراعى في كتابتها الحركات القصيرة والطَّوِيلة، مثل العديد من اللُّغات اللاتِينِيَّة.

١ , ٢– التَّحليل الصَّرفيّ (Morphological analysis)

ويتمُّ في هذه الدَّرجة تحليل الكلمة إلى عناصِرِها الأساسيَّة (السَّابق والجذع واللاَّحق)؛ ووفقًا لطبيعة اللُّغة العربيَّة الاشتقاقيَّة يوجد للكلمة العربيَّة – عادةً – أكثر من احتمال، إلاَّ أنَّنا نستطيع اختيارَ الاحتمال الأكثر مناسبةً من خلال السِّياق. ويتمُّ هذا الاختيار باستخدام الطُّرُق الإحصائيَّة الَّتي تُساعد – كذلك – في تحديد أقسام الكلام (Parts of Speech) وما يتعلَّقُ بها من توصيفات.

١ , ٣– التَّحليل النَّحويّ (Syntactic analysis)

ويتمُّ في هذا المستوى تعيين وظيفة الكلمة في الجُملة من حيثُ موقعُها الإعرابيّ، الأمر الَّذي يُساعد على فهم المعنى (الدِّلالة)؛ إلاَّ أنَّ الارتباطَ القائم بين التَّحليل النَّحوِيّ والمعنى يزيدُ الأمرَ صُعوبة في اللُّغة العربيَّة، نظرًا لمرونتها في ترتيب مُكوِّنات الجملة (كأن يتقدَّمَ الخبر على المبتدأ أو يتأخَّر الفاعل عن المفعول).

ومن ناحيةٍ أخرى، لا يمكن تجاهُل دور المعنى عند التَّحليل الصَّرفيّ في تحديد الوظيفة النَّحوِيَّة للكلمة. ومثال ذلك كلمة (عامل)؛ فهي من النَّاحية الصَّرفيَّة (اسمُ فاعل) من الفِعل (عَمِلَ)، لكنَّها من النَّاحية النَّحوِيَّة تصلُحُ فاعلاً أو مفعولا أو نعتًا أو غير ذلك.

١ , ٤ – التَّحليل الدِّلاليّ (Semantic analysis)

وللتَّحليل الدِّلاليّ درجاتٌ عديدةٌ، أدناها تحديد دلالة الكلمة في سياقِها (كأن نُحَدِّدَ دلالة كلمة «عَين» على الجارحة أو الجاسوس أو بئر الماء)؛ وقد أولَينا آليَّة التَّحليل الدِّلاليّ أهمِّيَّة خاصَّة – في مواضعَ مُتفرِّقة من فُصُول هذا الكتاب – نظرًا لأهمِّيتها مع أنها الأقل انجازًا.

وتجدرُ الإشارة إلى أنَّ التحليل الدِّلالي له العديد من الدَّرجات الفرعيَّة، مثل:

- فكّ الالتباس الدِّلاليّ للكلمة (Word sense Disambiguation – WSD)، ومثال ذلك كلمة «عين» حيث تأتي بمعنى الجارحة، أو بئر الماء، أو الجاسوس... إلخ. وكمبحث فرعيّ لدلالة الكلمة: التَّعرُّف على أسماء الكائنات (Named Entity Recognition)، ومن أمثلتها كلمة «القاهرة» الَّتي تأتي نعتًا لمؤَّنثٍ، أو علمًا على مدينة القاهرة.
- فكّ التباس المشترَك اللَّفظِيّ (Anaphora Disambiguation). ومن أمثلة ذلك: «أحبَّت البنتُ أختها كثيرًا، فلمَّا قابلتها جرت إليها». وموضع الالتباس أنَّنا نجد صُعُوبة في معرفة مَن الَّذي جرى إلى مَن، هل البنت إلى أختها؟ أم الأخت إلى البنت؟ ويجوز أن نسميه «إزالة التباس الضمائر».
- تحليل الإلماحات (Mention analysis). ومثال ذلك «قابل هشام محمدًا» فقال له: «موعدنا غدًا يا هشام» فرد قائلاً: «بل بعد غد». مَن قال ماذا؟
- فكّ الالتباس البلاغي (Rhetoric Disambiguation): ومثاله «رأيت أسودًا في المعركة»، إذ القصد «رأيتُ جنودًا شجعانا».
- الفصل بين الموضوعات (Subject separation). ففي أحيان كثيرة تتناول المقالة أكثر من موضوع ولا نجدُ بالضرورة عناوين فرعية للفصل بين الموضوعات الجزئية. نعم؛ يمكن أن يكون كل موضوع له فقرات منفصلة، ولكن عدد فقرات كل موضوع مختلف ولا يفصل بينها عنوان فرعيّ. فكيف يمكن أن نفصل بين الموضوعات الجزئية؟ إنَّ هذا الفصل يكون ضروريًّا عند استرجاع المعلومات المفيدة؛ فلو أمكن الإشارة إلى موضع المعلومة المفيدة في المقالة أو الكتاب بشكل دقيق، فإن ذلك يوفر على الباحثين الكثير من الوقت.

وليس ذلك فحسب. بل إنَّ الموضوعات المراد الفصل فيها تحت مسمى التحليل الدلالي كثيرة جدًّا؛ والمشكلة أن كلاً منها لم يبلغ مستوى حسمه - إن بدأ أصلا في العربية - مبلغًا مقنعًا ومفيدًا؛ إذ حين نصل إلى ٧٠٪ أو ٨٠٪ فإن هذا المستوى يُمَثِّلُ درجةَ حسم ضعيفةً للغاية لكثير من الغايات المطلوبة من الحاسوب.

١ , ٥- التَّحليل المُعجميّ (Lexical analysis)

ويتمُّ في هذا المستوى تعيينُ مُكوِّنات المُعجم اللُّغويّ وتوصيفها، على النَّحو الَّذي يُمكِّنُ من التَّمييز بينَ مباني اللُّغة ومعانيها. ويجمعُ هذا المُستوى بينَ مُستويات التَّحليل اللُّغويّ السَّابقة؛ حيثُ يُعنى بالتَّحليل الفونولوجي في معلومات نُطق المُفرَدات؛ ولهذا البُعد أهمِّيَّةٌ خاصَّة في المعاجم المُوجَّهة لغَير أبناء اللُّغة؛ ويُعنى بالتَّحليل الصَّرفيّ بهدف استخلاص الوحدات الأساسيَّة للمُعجم [المباني]؛ ويُعنى بالتَّحليل التَّركيبيّ في الاستدلال على سياقات المُفرَدات ومعانيها الوظيفيَّة؛ ويُعنى كذلكَ بالتَّحليل الدِّلاليّ في الاستدلال على المعاني المُعجميَّة.

أمَّا عن إشكاليَّة التَّداخُل بين هذه المستويات، فنستَطِيعُ أن نَتَبَيَّنَها من خلال النِّقاط التَّالية:

- يتداخل مستوى التَّحليل الفونولوجي مع مستوى التَّحليل الصَّرفيّ، حيث تقتَضِي طبيعة النِّظامُ الكتابيُّ للعربيَّة أن نجمعَ في مُستويات تحليل نُصُوصِها بين التَّحليل الفونولوجيّ والتَّحليل الصَّرفيّ. فمع وُجود ظاهرة الإعراب في العربيَّة، إلاَّ أنَّ أهلَها - عادةً - يُهملونَ تَشكِيلَ حُرُوفها بعلامات الضَّبط (الحركات القصيرة والشَّدَّة والتَّنوين) - باستثناء حالة التَّنوين بالفتح؛ ونُلاحِظُ أنَّ التَّشكيل لا يُزِيلُ لبس الصرف تماما، وإنَّما يُزيلُه الحسمُ الصَّرفيّ. ومثال ذلك: كلمة «قَالَ» بعد تشكيلها يظل الالتباس قائما، هل هي من جذر «ق و ل» من القَول، أم من جذر «ق ي ل» من «النوم وقت القيلولة». إذن يجوز أن نقول بلغة أهل الرياضيَّات: إن إزالة الالتباس الفونولوجي شرط ضروريّ، ولكنه غير كاف لإزالة الالتباس الصرفي؛ مع العلم بأن التحليل الفونولوجي معنيٌّ أيضًا بمسائل لها علاقة بنطق الكلمات؛ ويدخل في مباحثه الحروف غير المنطوقة والحروف المدغمة وما يحدث في حالات التقاء الساكنين... إلخ.

وكمثال للحروف غير المنطوقة، الألف بعدَ الواو في «خرجوا»؛ فإن الألف لا تنطق. ومثالٌ للحروف المدغَمة، لام التَّعريف في «الشَّجرة»؛ حيثُ تُدغَم اللام في الشين، وتُشَدَّد الشين بالتالي.

- يتداخل مستوى التَّحليل الفونولوجي مع مستوى التَّحليل النَّحويّ، ونلاحظُ هذا في الارتباط الوثيق بين التَّشكيل الكامل للكلمة العربيَّة والتَّوصِيف الإعرابيّ النَّاتج عن التَّحليل النَّحويّ للجُملة.

وعلى الرَّغم من انفصال العديد من الأبحاث في جانبي التَّحليل الدِّلاليّ والنَّحويّ، إلاَّ أنَّنا ندعو - عند التَّعامُل مع اللُّغة العربيَّة - إلى الرَّبط بين هذين المستويين رياضِيًّا، لأنَّ الأصل في النحو أن يكونَ فرع المعنى.

٢- التَّحدِّيات الَّتي تواجه تقنيات اللُّغات الإنسانِيَّة

اللغةُ أمانة عظيمة، ومقدرة الإنسان فيها غير قابلة للتقليد أو المحاكاة. فإذا جلست مع أخيك أو جارك وتسامرت معه لساعات طويلة فلا تكاد تمر - عبر آلاف الكلمات المتدفقة من فمه - كلمة لا تفهمها. فنسبة ما تفهمه إذا لم يكن ١٠٠٪ فإنه يتجاوز ٩ ,٩٩٪. بينما أفضل أنظمة التعرف على الكلام في الكلام التلقائي (كما في الاجتماعات) قد لا تتجاوز ٧٠٪، وقد تزداد إذا كان الكلام منتظما (كما في نشرات الأخبار)؛ فقد تصل في هذه الحالة إلى أكثر من ٩٠٪ بقليل). فإذا قارنَّا بين ٩ ,٩٩٪ وبين ٩٠٪ وجدنا للوهلة الأولى أن هذا شيء عظيم جدًّا وأنه لم يبق إلا القليل. وفي الحقيقة يتطلب تحقيق ٩٠٪ باستخدام الآلة سنوات من البحث والعمل، وقد لا ندرك بعد سنين طويلة من الأبحاث إلا زيادة ١٪ أو ٢٪.

أريدك أن تقارن بين نسبة الخطأ عند الإنسان وعند الآلة، فنسبة أخطاء فهم الكلام عند الإنسان في نفس الظروف أقل من ١ ,٠٪، أي أقلّ من كلمة من كل ١٠٠٠ كلمة، بينما يخطئ الحاسوب في فهم كلمة من كل ١٠ أو في أحسن الأحوال ٢٠ كلمة. لو أن الجملة العربية تحتوي على ٢٠ كلمة فإن الحاسوب سيخطئ في المتوسط في كلمة أو كلمتين في كل جملة وربما تكون هذه هي أهم كلمة في الجملة، أو ربما تعرف على الكلمة بكلمة أخرى تغير المعنى وتبدله.

الذي أود توضيحه أننا معذورون في عدم قدرتنا على اللحاق بمقدرة الإنسان ولو بعد حين.

في عام ٢٠٠٠م، في أحد المؤتمرات الدولية لأبحاث وتقنيات الصوت (Inter-Speech 2000, Istanbul-Turkey)، سأل أحد مدراء المؤتمر في ختامِهِ سؤالاً مفتوحًا وطلب من الحضور الإجابة؛ هل بعد ١٥ سنة من ذلك الوقت يمكن أن نعتمد على الحاسوب في فهم الكلام فهمًا يجعل الإنسان يعتمد عليه كليًّا؟ وكانَ إجماع شيوخ العلم وأساطينه باستحالة ذلك. فقد انقطع أمل العلماء الذين قضوا ٥٠ سنة أو تزيد في تقنيات الصوت في الوصول إلى شيء يقارب الإنسان، ولكن لم ينقطع الأمل في إنجاز مهام محددة، لكلماتٍ محدودة في مجالات تخصصية محددة؛ بحيث نخاطب الآلة (الحاسوب أو النَّقَّال) في جو هادئ وبشكل منتظم وبلغة صحيحة؛ أما أن نتعامل مع الآلة كما نتعامل مع بعضنا بعضًا، فهذا على قدر علمنا في هذا الزمن من المحال ولعل الله الذي تكرم بالبيان يتكرم بالعلم الذي يمكن الإنسان أكثر وأكثر.

حين قارن العلماء بين ما تم إنفاقه على أبحاث التعرف على الكلام وما تمَّ إنفاقه على أبحاث الأشعة في الطب في الثمانينيات من القرن العشرين، كان الإنفاق على الأبحاث متقاربًا وكانت النتائج أيضًا متقاربة، إذ قاربت النتائج في الحالتين ٩٠٪ كنسبة نجاح. ولكن شتان بين التأثيرين؛ ففي مجال الطب لا تكاد تجد مستشفى ولا مستوصفا إلا فيه جهاز أو أكثر للأشعة، وليس الحال كذلك بالنسبة للكلام، لأن ٩٩٪ هي بداية الرضا الحقيقي، وما دون ذلك يصعب انتشاره وقبوله، ناهيك عن ٩٠٪.

والسبب - كما ذكرنا سالفًا - أننا نملك أجهزة نطق وتحليل للكلام غير قابلة للمقارنة، فمن العسير جدا مضاهاة أذن الإنسان، حتى ٩٩٪ بها عدد من الأخطاء أكثر بعشر مرات من نسبته عند الإنسان (٩ , ٩٩ ٪)، والتي تكاد تكون مرجع قدرة الإنسان الراقية جدا في هذا المجال. سبحان من خلق فسوَّى وَقَدَّرَ فهدى.

ومع أننا استعرضنا في هذا الفصل تقنية التعرف على الكلام مع مقارنة قدرة الحاسوب مع قدرة الإنسان فإن كثيرًا من التقنيات يتفاوت رضا الإنسان عنه، بين أن يكون كاملاً أو ما يقارب الوضع مع تقنية التعرف على الكلام.

٣- التَّحدِّيات الَّتي تُواجه معالجة اللغة العربية

للُّغة العربيَّة العديد من السِّمات الَّتي تُمَيِّزُها، ومع أنَّ هذه السِّمات قد تُمَثِّلُ عناصِرَ قُوَّةٍ لها، إلاَّ أنَّها تُمَثِّلُ في الوقت ذاته عناصر تحدٍّ إذا ما أردنا إخضاعها للحوسبة.

وسنُورِدُ في هذا المحور بعضَ التَّحدِّيات الَّتي تُواجهُنا عند إخضاع العربيَّة للمُعالجَة الحاسوبيَّة - من خلال النِّقاط التَّالية:

٣,١- الكتابة العربية

- الكتابة العربية لا تُفهم إلا إذا كانت حروفها متصلة، كما في هذا الكتاب:
  - هناك بعضُ الحروف الَّتي لا تقبل الاتِّصال، إذ بينها وبين مُتجاوراتِها من الحُرُوف في نفس الكلمة مسافة قصيرة نسبيًّا. والواقعُ أنَّ فصلَ الكلمات بعضها عن بعض مسألة ليست بسيطة. ففي أنظمة تجارية لبرامج التعرف على الحروف العربية نجدُ أنَّ نسبة تقارب ١٠٪ من الأخطاء على مستوى الكلمات، يكونُ أصل الخطأ فيها خطأ في فصل الكلمات بعضها عن بعض، هذا اذا كانت الوثيقة واضحة.
  - الكلمة العربية - والتي يجب أن تكتب فيها الكثير من الحروف متصلة ببعضها - تجعل مسألة التعرف على حروف هذه الكلمات مسألة أكثر صعوبة من التعرف على النصوص المكتوبة في اللغات اللاتينية، حيث يمكن أن تكتب جميع حروفها بطريقة مُنفصلة عن بعضها البعض (وهو الغالب في الطباعة). وذلك يضيف تحديًا آخر أمام تقنية تحويل الكلام العربي المطبوع إلى نصوص مقروءة. وسوف نتعرف على ذلك بالتفصيل في الباب الخامس عشر من هذا الكتاب.
- أسماء الأعلام والكائنات (Named Entities) لا تُمَيَّزُ في اللغة العربية، بينما في بعض اللغات تميز بحروف كبيرة في الرسم (Capital letters). مثال ذلك «زرت القاهرة» مقابلها (I have visited Cairo). فإن وجود هذه العلامة الإضافية في الكتابة ترفع نسبة التعرف على أسماء الكائِنات آليًّا (Name Entity Recognition)، وهذه كثيرًا ما تستدعي في مجال محركات البحث، إذ لا يكون

الاستفسار عن الكلمات العادية، ولكن عن مُكوِّنات اسميَّة سواء أكانت أشخاصا أم أحداثا أم مؤسسات....إلخ.

- غياب التشكيل عند بنية الكلمة أو نهايتها. فقد اعتادَ العرب في مُعظَم كتاباتهم عدم وجود تشكيل على الحروف، وإذا وجد (ربما في ١٠٪ من الكتب) يكون تشكيلاً جزئيًّا لبيان الفعل المبني للمجهول من الفعل المبني للمعلوم مثلاً، أو لبيان التشديد في الكلمات التي يغير فيها التشديد المعني... إلخ. وهذا يدع لنا الكلمة العربية من وجهة النظر الحاسوبية لغزًا كبيرًا؛ فكلمة مثل عين يمكن أن تكون:- عَيْن أو عَيَّنَ أو عُيِّنَ ..الخ. التشكيل في الحقيقة يمثل جزءً ضروريًّا من حقيقة الكلمة العربية، إذ إنه يمثل الحركات القصيرة (Short vowels)، والَّتي عادة ما تكتب في لغات أخرى كالإنجليزية والفرنسية،... إلخ. إذ يترك للقارئ العربي استنباط ما لم يكتب من هذه الحركات أو التشديد أو التنوين لبداهته، وإذا كان العقل البشري قادرًا على الحسم بيسر وسهولة في قريب من ١٠٠٪ من الحالات، فإن الأمر بالنسبة للحاسوب ليس بنفس السهولة.

هذا الأمر يسهل الكتابة العربية؛ فمتوسط عدد حروف الكلمة العربية المكتوبة يزيد قليلا من ٤ حروف؛ بينما يزيد الأمر في الإنجليزية مثلاً عن ذلك كثيرًا (نظرا لكتابة كل شيء تقريبا، باستثناء «الكلمات الوظِيفِيَّة function words»)، ولكنه يزيد دائرة الالتباس عند الحاسوب ويجعل المعالجة الحاسوبية ذات أبعاد أكثر من مثيلاتها في بعض اللغات اللاتينية. ومن نعمة الله أن أصبَحَت الكتابةُ العربيَّة منقوطة؛ فلو بَقِيَت الحُرُوفُ العَرِبيَّة دونَ تنقيطٍ لأصبَحَت مُهِمَّةُ الحاسوب شبهَ مُستَحيلة.

٢,٣- الأخطاء الشائعة

كثيرة في كل اللغات ولكن في اللغة العربية هناك حالات تتسبب في زيادة نسبة هذه الأخطاء، منها:

- الهمزات، خاصة مع الألف (ا، أ، إ)؛ فنسبة الأخطاء فيها عالية بين عامة الناس.
- الهاء والتاء المربوطتان؛ مثال «القاهرة» كثيرًا ما تُكتب «القاهره» والعكس «ضربه» يمكن أن تكتب «ضربة».

- قد لا يكون خطأ ولكن اعتاد أهل مصر مثلا عدم استعمال الياء آخر الكلمة منقوطة لتمييزها عن الألف المقصورة كما في «على» و «علي».

هذه الأخطاء الشائعة تضيف للالتباس تحديًا آخر، إذ إنَّ أيَّ نظام لتحليل النص العربي الخام مثلا أو التشكيل يتطلَّب - لكي يكون مقبولاً عمليًّا - التعامل مع النصوص العربية المرتبكة في همزاتها أو هائها وتائها أو يائها وألفها المقصورة.

- نظرًا لدخول الثقافة الأجنبية بشدة وشيوع الكتابة بها، فإن كثيرًا من النصوص المطلوب التعامل معها على الإنترنت مكتوبة بالحروف اللاتينية. ومثال ذلك: (Dadi ana bahebk a'wy) ويقصد الكاتب: «أبى أنا أحبك كثيرًا».
- نظرًا لإهمال التعامل بالعربيَّة في المرحلة الجامعية في معظم الجامعات العربيَّة، نجد الكثيرين من خريجي الجامعات تختلط العامية في كتاباتهم بالعربية مثال: «معلهش أنا كنت فاكر إني هاقدر أجي امبارح ...».

وأحيانًا نجد أننا نود معرفة آرائهم وأفكارهم فنجد أنه لا بد من محاولة ترجمة مقالاتهم كما هي إلى العربية الفصحى الحديثة (Modern Standard Arabic– MSA).

- الأخطاء الإملائيَّة الشَّائعة، مثل:

شَيْء ← (شئ - شيئ).

إن شاء الله ← (إنشاء الله).

لا بُدَّ ← (لابد).

- ناهيك عن أخطاء الهجاء الأخرى، كمن يستعمل الحروف القريبة في النطق فيكتب:

مظاهر ← مضاهر (وهو خلط شائع في الجزيرة العربية والمغرب العربيّ).

ذليل ← زليل.

وسيط ← وصيط ... إلخ.

إلا أن هذا النَّوعَ منتشر في كل اللغات الإنسانيَّة ويكثر بين متعلمي العربية كلغة ثانية.

٣,٣- التركيب الصرفي للكلمة العربية

الكلمة العربية مركبة تركيبًا عميقًا مما يجعلها من اللغات ذات المفردات الكثيرة (morphologically rich language) مقارنة بالإنجليزية مثلاً، وسوف نعرض في فصول الكتاب لتركيب الكلمة العربية. ولكن ما يهمنا هنا الإشارة إلى حقيقة عدد الكلمات العربية والتي يمكن أن تكون بالملايين، ومع أن الكلمة العربية تتكون من عدد محدود جدا من الوحدات المعجَميَّة (Lexemes) (~ ٥٠٠٠ جذر، ~ ١٠٠ وزن (عدا ما يتفرع منها نتيجة الاعتلال)، ~ ٣٠٠ سابق و ~ ٥٥٠ لاحق). وهـذا في مجموعه لا يتجاوز ٦٠٠٠ وحدة مُعجَمِيَّة (Lexemes)، فهذه ميزة كبيرة جدًّا لم تستفد منها كثير من التقنيات بالقدر الكافي بعد. وما الصعوبة التي يشكلها العدد الكبير من الكلمات في اللغة؟ على سبيل المثال: التعرف على الكلام، هذه التقنية تتأثر نتائجها بعدد الكلمات التي تتعامل معها في أي مهمة من مهام هذه التقنية. وللعلم، فإنه في مجال الأعمال يكفي في الإنجليزية نحو ٦٤ ألف كلمة لتغطي ٩٩٪ من احتياج المتحدث بينما يحتاج العربي إلى أكثر من ٦٠٠ ألف كلمة ليغطي ٩٩٪ من احتياجاته في نفس المجال.

٣,٤- تركيب الجملة العربية

- تسمح اللغة العربية بالتقديم والتأخير بين كلماتها، ولهذه الظاهرة دلالات بلاغية وقيمة رائعة. فمثال ذلك: قوله تعالى ﴿إِيَّاكَ نَعْبُدُ وَإِيَّاكَ نَسْتَعِينُ﴾ [الفاتحة ٥]. حيثُ يُفيدُ تقديم «إياك» على «نعبد» الحصر والقصر. وحين يقول ربُّنا - سبحانه وتعالى ﴿يُريدُ الله بكم اليُسْر﴾ [البقرة ١٨٥]. ﴿الله يُريدُ أن يتوبَ عليكم﴾ [النِّساء ٢٧]؛ فقد اختلفَت دلالة التَّقديم والتَّأخير بين الفعل والفاعل من آيةٍ إلى أخرى.

ولكن ما تأثير ذلك على الحوسبة اللغوية؟ إنه يزيد من صعوبة المعالجة النحوية والبلاغية للعربية مقارنة بالإنجليزية مثلاً، وفي نفس الوقت لم يقدم لنا اللغويون مجموعة متكاملة من القواعد التي تحيط بالموضوع وتحسمه إذ إن كثيرًا من الحسم يعتمد على الجانب الدلالي، على اعتبار أن النحو فرع المعنى.

ولا بد من صياغة القواعد اللغوية بطريقة يمكن الاستفادة منها حاسوبيًّا. ونضرب لذلك مثالاً:

تتبع الصفة الموصوف في العدد والجنس والتعريف (مع وجود أنواع أخرى للصفة). وعليه، فالقياس الحاسوبي للعدد والجنس والتعريف ممكن، وبذلك تنجح هذه القاعدة في البرمجة الحاسوبية. بينما إذا قلنا إن الخبر هو الذي يُتمُّ معنى المبتدأ فإننا ربطنا الحسم النحـــوي بالدلالة وضـــاع طريـــق الحسـم من تحت أقدامنا. إذ إن الدلالة أعلى في سلم الحسم اللغوي من النحو.

- غياب بعض الكلمات تمامًا مع وجودها التقديري: ومثال ذلك أنك تتحدث عن «زيد» من الناس ثم تقول: (ودخل إلى الحديقة). أين الفاعل الذي دخل؟ تقول إنه لضمير مستتر تقديره هو. قد تكون هذه الظاهرة منتشرة في لغات أخرى بيد أنها تزيد التحدي أمام المعالجة الآلية.
- يمكن أن يغيبَ الفعلُ عن الجملة العربية جملة وتفصيلاً وتسمى جملة اسمية، بينما الجملة الإنجليزية لا بد من وجود فعل بها، وإن اقتضى الأمر استخدم فعل مساعد: (auxiliary verb: verb to be).

«الجو جميل» المقابل لها بالإنجليزية (The weather is beautiful).

إن ذلك يرفع العربية درجة في مرونة استخدامها ويجعلها أرفع وأيسر في الاستخدام البشري، ولكنه يوسع دائرة الخيارات غير المحسومة أمامنا عندَ إخضاع الآلة لهذا العمل.

- ومن التحديات التي تواجهنا أن كاتب العربية يستطيع أن يربط بين جملتين بحرف الواو، إذ إن كتابته أسهل عليه من كتابة فاصلة أو نقطة؛ لذلك فمتوسط طول الجملة العربية أكبر بكثير من متوسط طول الجملة الإنجليزية. وماذا يشكل هذا بالنسبة لحوسبة اللغة؟

إن طول الجملة يصعب على الحاسوب فهمها فهمًا دقيقا. خذ هذا المثال؛ لو أن تقنية وصلت لدقة ٩٠٪ في المتوسط لمعالجة كل كلمة على حِدة، ستكون نسبة الدقة التي لجملة من:[1]

- كلمتين، دقة الجملة = (دقة الكلمة)٢ = (٩٠ ٪)^{٢} = ٨١ ٪
- ٥ كلمات، دقة الجملة = (دقة الكلمة)٥ = (٩٠ ٪)^{٥} = ٥٩ ٪
- ١٠٠ كلمة، دقة الجملة = (دقة الكلمة)١٠٠ = ٠,٠٠٢٧ ٪ أي تقريبًا صفر ٪؛ وهذا يعني أنه كلما طالت الجملة كلما صعبت معالجتها.

٣,٥- التعامل مع الواقع اللغوي

فالعربيَّة تُعاني تراجُعًا كبيرًا في السنوات العشرين الأخيرة، وبالتحديد بعد انتشار الإنترنت وزيادة نسبة التعليم باللغات الأجنبية في مراحل التعليم قبل الجامعي؛ فنحن الآن أمام ظواهر لغوية خطيرة تهدد مستقبل العربية في أرضها. فما الموقف من العامية والكتابة بالحروف اللاتينية؟

- لسنا من أنصار تمزيق أهل العربية باللهجات، ونتمنى أن تجمعهم الفصحى، على الأقل فهمًا وكتابة. بمعنى أنَّنا لا نريد أن تنتشر أي قواعد للكتابة والقراءة إلا قواعد الفصحى.
- إذا كانت هناك بعض اللهجات، فإننا ندعو إلى التَّقريب بينها؛ وربما يجعلُ عالم الإنترنت ذلك ممكنا. ولكن كما سمح الله تعالى بتلاوة كتابه بقراءات مختلفة لاختلاف اللهجات، فإنَّ علينا أن نقبلَ بوُجود لهجاتٍ مُختلفة، ولكن.. ليكن ذلك في القراءة دون الكتابة. فمثلاً أبحاث نطق النصوص العربية (Arabic Text to Speech) لا بد أن تنطق النصوص بالفصحى لا العامية؛ والفصحى يفهمها عامة العرب، وذلك لأن نشرات الأخبار في معظم القنوات تنطق بالفصحى (حتى مع بعض الأخطاء).

---

١ - يمكن تحديد دقَّة الجُملة من خلال نسبة الجُمل الَّتى تكون جميع كلماتها صحيحة.

أما بالنسبة للتعرف على الكلام من العامة فإننا نجد لزامًا على الباحثين أن يتعاملوا مع الواقع لخدمته وفهمه. فتطوير نظام للتعرف على الكلام بالعامية قد يكون مهما لخدمة عامة الناس في شؤون حياتهم المختلفة.

٤- التَّعريف بتقنيات اللُّغات الإنسانيَّة

كانَ للتَّقدُّم المعلوماتيّ الَّذي شهده العالَمَ في العُقود الأخيرة أثرٌ كبيرٌ في تطوير أدوات مُعالَجَة اللُّغات الحيَّة وتنمية تطبيقاتها الَّتي تزدادُ يومًا بعد يوم. وتتفاوت درجات العناية بالتِّقنيات المستَخدمة في بناء وتطوير هذه الأدوات بحسب الحاجة إليها؛ كما تتنوَّعُ هذه التِّقنيات بتنوُّع بيئات العمل فيها. وفي ضوء هذا، يُمكن تقسيمُ تقنيات اللُّغات الحيَّة إلى الأقسام الموَضَّحة في الشَّكل (١-٢).

مِن تِقنيات اللُّغات الإنسانِيَّة

تِقنيات مُرتَبِطة بالنُّصُوص	تِقنيات مُرتَبِطة بالصَّوت	تِقنيات مُرتَبِطة بالكتابة	تِقنيات أخرى مركبة
البحث في النُّصُوص	التَّعرُّف الآليّ على الكلام	التَّعرُّف على الكتابة المطبوعة	تقنية قراءة الكُتُب
التَّرجمة الآليَّة	تحويل النَّصّ إلى كلام	التَّعرُّف على الكتابة المخطوطة	التَّرجمة الشَّفَهيَّة الآليَّة
الإجابة الآليَّة عن الأسئلة	التَّعرُّف على اللُّغة	التَّعرُّف الآنيّ على الحُرُوف	الاستِجابة الصَّوتيَّة التَّفاعُليَّة
تقييم الكتابة الإنشائيَّة	تعلُّم النُّطق بمُساعدة الحاسُوب	تعلُّم الخطّ بمُساعدة الحاسُوب	تحويل لغة الإشارة إلى كلام والعكس
التَّقييم الآليّ للنُّصُوص			أنظمة إدارة الحوار
تصنيف وتجميع الوثائق			الأنظمة المتَقَدِّمة لاسترجاع المعلومات
التَّنقيب عن الآراء			التَّنقيب في النُّصُوص
المدَقِّقات اللُّغَويَّة			
تعلُّم اللُّغات بمُساعدة الحاسُوب			
القِياس الأسلوبيّ			

الشَّكل ١-٢: من تقنيات اللغات الإنسانيَّة.

وسنُحاولُ - فيما يلي- أن نعرِضَ - بإيجازٍ- لهذه التِّقنيات، على أن نعود لتفصِيل الحديث في المشهور منها في فُصُول الكتاب اللاحقة.

٤ , ١ - البحث في النصوص «محركات البحث» (Search Engines)

تعد محركات البحث في النصوص أو الوثائق من أكثر التقنيات استخداما، والغرض منها استخراج المعلومات المطلوبة للمستخدم، ولعلَّ المحرِّكات الأكثر استخداما في الوقت الحالي هي تلك التي تُعنى بالبحث عن الوثائق التي تحتوى على كلمات بعينها وتسمى الكلمات المفتاحية. ويمكن تقسيم محركات البحث إلى عدة أنواع، منها:

- مُحرِّكات البحث في الوثائق الشَّخصيَّة على الحاسوب الشخصيّ.
- مُحرِّكات البحث في الوثائق المؤسسيَّة (Enterprise Search).
- مُحرِّكات البحث في الوثائق الشَّبَكِيَّة (البحـث عن وثائق الإنترنت) (web Search).

ويمكن تقسيمها من حيث نوع النصوص إلى:

- بحث عام.
- وبحث في مجال متخصص، مثل الطب، والقانون،....إلخ.

وتجدر الإشارة إلى أن البحث في المجالات المتخصصة يحتاج إلى ما يُسَمَّى بالبحث الدلالي أو البحث الأنطولوجي، ويُساعد هذا البحث على دقة وسرعة الوصول للمعلومة. وعلى سبيل المثال: لو أردنا أن نبحث عن كلمة «وحوش»، لن تأتي بالوثائق التي تتحدث عن الأسود، أو النمور، أو الضباع،... إلخ. فهذا يعد قصورا في البحث، ما لم ندعم محركات البحث بمعلومات دلالية لكل كلمة؛ ولا يزال البحث الدِّلاليّ قاصرًا ونادرًا في اللُّغة العربيَّة.

كما ظهر - حديثًا - نوعٌ من البحث، يتجاوز النصوص إلى الوسائط المختَلِفة الأخرى، مثل:

- البحث في ملفات الصوت.

- البحث عن الصور.
- البحث في ملفات المرئِيَّات.

٤ , ٢- التَّرجمة الآليَّة (Machine Translation)

تعد الترجمة الآلية من التقنيات المهمة، والتي تضاعفت قيمتها بعد ظهور الإنترنت ودعت الحاجة للتواصل بلغات مختلفة أو للاستفادة من كنوز معرفية هائلة ولكن بلغات أخرى. وللترجمة الآلية مدارس عدة منها:

- مدرسة الترجمة بقواعد وقواميس ثنائية اللغة.
- مدرسة الترجمة بالاعتماد على مدونة ثنائية اللغة، حيث يوجد في هذا النوع من المدَوَّنات ما يكفي من معلومات لتعليم نماذج رياضية. وهناك طريقتان لهذه المدرسة:
  - طريقة التعلم من الأمثلة؛ وتستخدم إذا قلَّ حجمُ المدونة الثنائيَّة اللغة.
  - طريقة الترجمة الآلية الإحصائية (Statistical Machine Translation)؛ وهذه تحتاج لمدونات كبيرة جدا قد تصل إلى عدة ملايين من الجمل لكي تعطي نتائج ملموسة. وتعد هذه الطريقة هي الأكثر شيوعا بين العاملين في حقل الترجمة حديثا.
- مدرسة الترجمة عبر لغة وسيطة، وقد حدث تطور في هذه المدرسة أيضا ولها بعض الإنجاز الجدير بالمراجعة والاستفادة منه. وقد خصص باب كامل لدراسة هذا الموضوع بالتفصيل في الجزء الثاني من الكتاب.

وقد استخدم علماء هذا التخصص عدة وسائل لقياس جودة الترجمة الآلية؛ لعلَّ أشهرها ما يُعرَفُ بـ: مُقَيِّم «بلو» أو (BLEU Score) حيث يسجل الإنسان - إذا قامَ بالتَّرجمة بنفسِه- أرقاما تصل إلى +٨٠٪ على هذا المقياس، بينما تسجل أفضل محركات الترجمة الآليَّة من اللغة العربية للإنجليزية -٥٠٪. ولازال التقدم متلاحقا، والاهتمام العلمي في تزايُدٍ لتحقيق هدف البشرية في تحقيق التواصل بين البشر وتوسعة الاستفادة من معارفهم المختلفة للغات. وتتيح محركات الترجمة الآلية الترجمة من وإلى العديد من

لغات العالم. وعلى سبيل المثال: يتيح محرك الترجمة الآلية في جوجل الترجمة بين أكثر من ستين لغة.

٤ ,٣- الإجابة الآلية عن الأسئلة (Question Answering)

تعد الوسيلة الأولى لاكتساب المعرفة هي السؤال في صيغة اللغة الإنسانية التي يتحدثها الإنسان؛ والمطلوب تحليل السؤال وفهم المراد والبحث عن الإجابات بين الوثائق المتاحة على الإنترنت أو في قواعد بيانات مؤسسة من المؤسسات، ثم عمل استخلاص للمعلومات بشكل كافٍ للإجابة عن السؤال ثم وضع هذه الإجابة في صيغة لغة السائل الإنسانية. وهذه طريقة مركبة وتحتاج إلى عدد من التقنيات للوصول بالإجابة إلى درجة الدِّقة المناسبة. ومثل الكثير من تقنيات اللغات الإنسانية، فإن هذه التقنية المركبة لا تزال تحتاج إلى مزيد من الجهد وخاصة للغة العربية. ولا تتجاوز الأسئلة – في أكثر الأنظمة الخاصَّة بهذه التّقنية – نطاقًا محدودًا من المعارف.

٤ ,٤- التَّقييم الآليّ للمقال (Automated Essay Scoring)

نظرًا لصُعوبة تقييم الامتحانات والواجبات في مراحل التَّعليم المختلِفة على القائِمِين بها، فقد ظهرت أنماطٌ جديدةٌ من الأسئلة الَّتي يستطيع الحاسوب أن يتعاملَ معها بسهولة، وهي الأسئلة التي لها إجابات محدَّدة (مثل: الاختيار من متعدد، تعيين الصَّواب والخطأ، ملء الفراغات، وغير ذلك). ومن السَّهل إخضاع الحاسوب لتقييم هذا النَّوع من الأسئِلة بعد إمداده بالإجابات الصَّحيحة للأسئلة المدرَجة..

إلا أن هذه الأنماط من الأسئلة لا تختبرُ مهارات التعبير عند الطلاب؛ لهذا.. ظهرت الحاجة إلى تطوير تقنية يمكنها التعامل مع الإجابات الإنشائِيَّة (كموضوعات التَّعبير وما شابَهَها من أشكال الكتابة الإنشائِيَّة)؛ وظهرت بالفعل العديد من الامتحانات الدَّولية الَّتي تُفِيدُ من هذه التقنية.

وتقومُ فكرة التقنية – عُمومًا – على اختيار نماذج وعينات من أشكال الإجابة الإنشائِيَّة (الجيِّدة والمتوسِّطة والرَّدِيئة) وتصنيفها بحسب درجة التَّقييم وأخذ ملامحها والاحتفاظ بها، ثُمَّ تعيين الملامح الَّتي تُمَيِّزُ الإجابات الجديدة تمهيدًا لقياس درجة اقتراب الإجابة الجديدة من الإجابات المدرَجة في نظام التَّقييم. وتكمن الصُّعُوبة الحقيقيَّة في

كيفيَّة تعيين الملامح الممَيِّزة للإجابات الجديدة وطريقة قياسها بإجابات النِّظام.

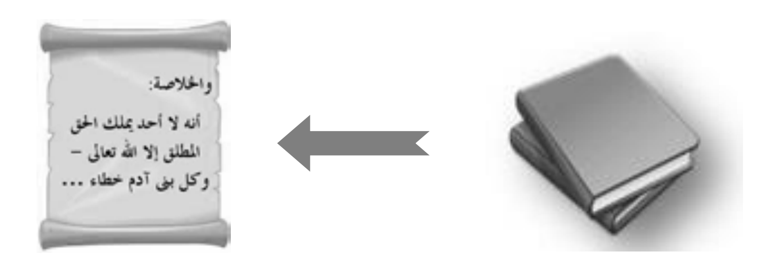
٤,٥- التَّلخيص الآليّ للنُّصُوص (Automatic text summarization)

دعت الحاجة العملية إلى هذه التقنية وتزايدت بعد الانفجار المعرفي الحالي والمنتظر زيادته مستقبلا. فإذا أمكن التعرف على أهم ما ورد في ورقة بحثية أو كتاب في موضوع ما، قبل استقطاع وقت كبير لقراءته بالكامل، فربما يفي الجزء المركز والمختصر بالغرض أو يقنع المستخدم لهذه التقنية بقراءة البحث أو الكتاب كاملا لبلوغ الهدف.

ولهذه التقنية طريقتان:

- الطَّريقة الأولى: الاستخلاص (Extraction)، وتعتمد على انتقاء الجمل الأهم من الوثيقة.
- الطَّريقة الأخرى: التَّلخيص (Abstraction)، وتعتمد على صياغة جديدة مختصرة لنفس المفاهيم.

والطَّريقة الأولى هي الأكثر شيوعًا واستخدامًا. أما الطَّريقة الأخرى فإن نجاحاتها في نطاقات محددة، ويزداد الاعتماد عليها كلما تقدمت نماذج فهم اللغات الإنسانية. ويدعو الطموح والحاجة العلمية والثقافية لدى الإنسان إلى الرَّغبة في أن يتمكن يوما ما من تلخيص مجموعة من الوثائق في مجال معين وبعض هذه الوثائق بلغات أخرى، بحيث تقوم محركات التلخيص بتلخيصها جميعا ووضع الخلاصة في بضعة أسطر أو صفحات للمستخدم. لا شك أن هذا أمل يضاعف من استفادة البشر من محيطات المعارف الحديثة ويزيد من إنجازهم، ولكن تحقيقه يحتاج لجهود كبيرة جدًّا لنتمكَّن من الاقتراب منه.



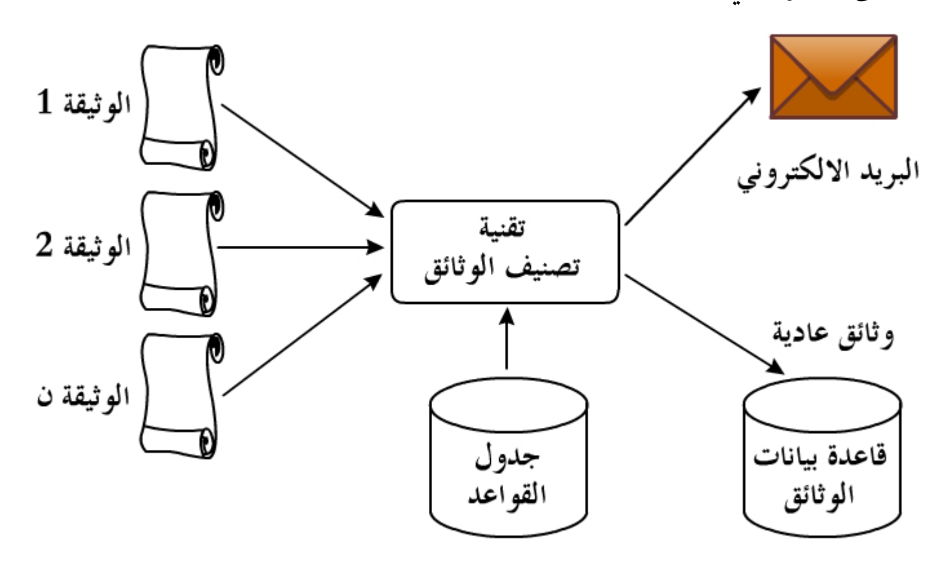
الشَّكل ١-٣: التَّلخيص الآليّ للنُّصُوص.

٤,٦- تصنيف وتجميع الوثائق
(Document Classification & Clustering)

عند استخدام محركات البحث للوصول إلى وثائق مهمة فإن الإنسان يحتاج إلى الوثائق المشابهة لما اختاره. وذلك يمكن أن يتم بطريقتين:

- تصنيف الوثائق آليا (Document Classification): حيث يكون هناك تقسيم مسبق للوثائق (كأن تقسم إلى سياسة واقتصاد...الخ) وهناك تصنيفات متعارف عليها لدى علم المكتبات، وكذلك تصنيفات فرعية لهذه التصنيفات.
- تجميع الوثائق الآلي (Clustering Document): وفي هذه الحالة ليس لدينا تقسيم مسبق للوثائق. وتُستَخدَم هذه الطَّريقة إذا كانت لدينا مجموعة من النصوص أو الوثائق وأردنا عمل تجميعات لأنواع الوثائق المتشابهة.

والهدف في الحالتين أن نصل إلى الوثائق المتشابهة. ويتمُّ تصنيف الوثائق وتجميعها آليًّا على النَّحو التَّالي:



الشَّكل ١-٤: تصنيف وتجميع الوثائق.

في حالة التصنيف الآلي: تُؤخذ ملامح كل وثيقة وتُحَدَّدُ خصائصها (في شكل رياضي) ويتم حساب درجة قرب أو بعد الوثائق من بعضها البعض حسب الملامح،

ويتم حساب المسافة بين الوثيقة المراد تصنيفها والوثائق المصنفة سلفا (سياسة، اقتصاد،... إلخ)، فإلى أي قسم كانت الوثيقة أقرب تصنف على هذا القسم.

وفي حالة تجميع النصوص آليًّا إلى مجموعات: يتم حساب قرب الوثائق كلها من بعضها البعض، وكل مجموعة منهم تقاربت ملامحها يتم ضمها في مجموعة مستقلة.

٤ , ٧- التَّنقيب عن الآراء (Opinion Mining)

من المهمّ أن نتعرَّفَ على اتجاهات الرأي على مستويات عدة؛ على المستوى السياسي لمعرفة ميول الناس إلى أي حزب أو فرد، وعلى المستوى الاقتصادي، كمعرفة رأي الناس في منتج جديد، إذ يكون مفيدا للشركة المنتجة أن تعرف رأى المستخدمين له. ولكن كيف سيتم تجميع هذه الآراء؟ تقليديا يتم عمل استبيانات وتُختار العينة أو العينات بعناية لتمثل رأي القطاع المستهدف من الدراسة....إلخ.

وتكون عملية اختيار العينة ثم إقناع أفرادها بملء هذه الاستبيانات مجهدة ومكلفة جدا. وأصبحت هناك مؤسسات متخصصة في هذا الأمر. ولكن بعد ظهور الإنترنت والمدونات (المقالات التي يسمح للآخرين بالتفاعل معها ووضع آرائهم) وشبكات التواصل الاجتماعي (Social networking) التي يُعَبِّر الناس فيها بحرية كاملة عن آرائهم، حتى في البلدان التي لا تكون فيها الحرية كاملة للناس، لأنه يمكن دائما استخدام أسماء مستعارة والهروب من الرقابة، أصبحت آراء الناس في كل شيء تقريبا متاحة وبكميات هائلة على الإنترنت. لذلك بدأت الحاجة لهذه التقنية والتي يكون من مكوناتها تجميع المواد ذات العلاقة بموضوع الرأي المراد معرفة اتجاهاته ثم اختبار هذه الآراء والخروج باتجاهاتها.

مثال: لو أن شركة سيارات أنتجت موديلاً جديدًا من السيارات وتريد معرفة آراء الناس فيه، فبعد تجميع المواد المنشورة على الإنترنت حول هذا الموديل من السيارات يمكن للشركة معرفة رأي الناس في السعر والجودة وما يعجبهم وما لا يعجبهم من صفات وخصائص السيارة الجديدة، ويمكن لهذه التقنية أن تستخدم وبشكل مستمر لإعطاء مطوري الأعمال في هذه الشركة معلومات قيمة يمكنهم من خلالها تحسين وتجويد منتجهم بشكل مستمر.

ويتم بعـــــد تجميع المواد عمل تصنيف للآراء حسـب الكلمات المستخدمة. فكلمات مثل: (رائعة، وجميلة، ومدهشة،...) تعبر عـن آراء إيجابيـة، وكلمات مثل: (سيئة، غاليـة جدا،...) تعبر عن آراء سلبية. طبعا هناك تعقيدات كثيرة للوصول الدقيق للآراء.

فعلى سبيل المثال: ماذا لو كتب الناس (ليست سيئة، أو ليست جميلة)؛ فورود كلمة (جميلة) مسبوقة بالنَّفي تقلب الإشارة الإيجابيَّة في الكلمة إلى إشارةٍ سلبيَّة. كل ذلك يجب أن يؤخذ في الاعتبار.

٤ ,٨- المدقق الإملائيّ، والنحويّ، والدلاليّ
(Spelling, Syntax and Semantic Checker)

بعد ظهور الحاسوب الآلي الشخصي وانتشاره في الثمانينيَّات من القرن الماضي انتشرت برامج الكتابة عليه مثل: برنامج «MS-Word©» من ميكروسوفت. وقد دَعَت الحاجة إلى ظهور تقنيات تكتشف الأخطاء الإملائية والنَّحوِيَّة والدِّلاليَّة في النُّصُوص، وتساعد في كثير من الأحيان على تصويب آلي للكلمات الخاطئة.

وهناك عدَّة طرق لإعداد هذه التقنية، منها:

- عمل قائمة بالكلمات الصحيحة، وعرض كل كلمة على هذه القائمة، فإن كانت الكلمة المكتوبة تتطابق مع إحدى كلمات القائمة فهذا يعني أنها صحيحة، وإلا فإنها كلمة خاطئة ومن ثم ننتقل للخطوة الثانية.
- نبحث عن أقرب كلمات إلى الكلمة التي تبيَّنَ خطأ كتابتها إملائيا، فنحضر قائمة موجزة من عدد محدود من الكلمات، ثم يتم اختيار أقرب الكلمات باختبار السياق (مجموعة الكلمات في الجوار) وأيُّها كان لها أعلى احتمال تم اختيارها بدلاً من الكلمة الخطأ.

والتقدم في تقنية التصحيح الإملائي عالٍ ومُرضٍ إلى حد كبير، لكنَّ الأمر ليسَ بسيطًا دائمًا؛ فماذا لو كُتِبَت كلمة صحيحة، لكنَّها غير مطلوبة في السِّياق؟ لو أرادَ الكاتب أن يقول: «كَلَّ ساعده» فكَتَبَ «قَلَّ ساعده».. نُلاحظ أنَّ كلمة «قلَّ» صحيحة إملائِيًّا، ولكنَّها ليسَت صوابًا في السِّياق. إنَّ اكتشافَ هذا النَّوع من الأخطاء يكون أكثر صُعوبة. ولكن لا يزال أمام تقنيتي المصحح النحوي والدلالي بعض الوقت للنضج

وعموم الاستخدام بالنسبة للغة العربية، على أن هناك مصححات نحوية متاحة للغات أخرى مثل الإنجليزية. المصحح الدلالي عموما أصعبهم ولم يصل لدرجة النضج الكافية بعد إلا في تخصصات محددة.

٤ , ٩- تعلُّم اللُّغات بمساعدة الحاسوب (Computer-Aided Language Learning)

إن تعلم اللغات فرع من العلوم يحوز على أكثر من ٢٠٪ من الجهد التعليمي للتعليم قبل الجامعي، وينقسم تعلم اللغات إلى تعلم اللغة الأم وتعلم لغات أخرى كلغات ثانية، إذ إن أخطاء كل منها ومُشكلات تعلمه مختلفة. وعموما تتعامل تقنيات المساعدة في تعلم اللغات في اكتشاف وتصويب أخطاء:

- الكتابة الإملائية.
- النحو وقواعده.
- الدلالة (استخدام اللفظ المناسب في المكان المناسب).

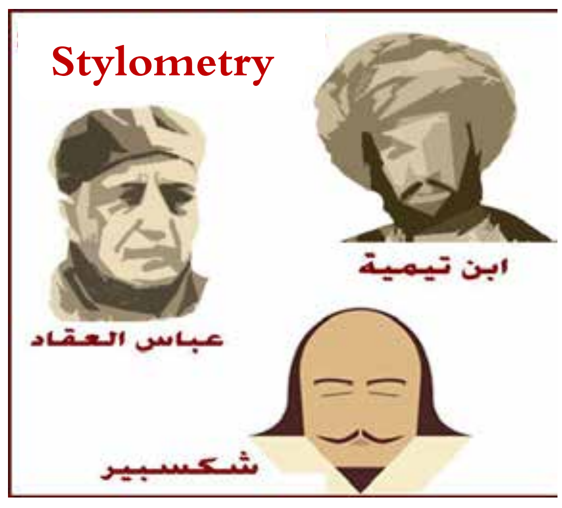
كما أن هناك تقنيات أخرى مفيدة في مجال تعلم اللغات لها علاقة بتصويب النُّطق وتصويب الكتابة ... إلخ.

وكما أسلفنا في الحديث عن المصحح الإملائي، تُبنى التقنيات على مثل ذاك المنهج. والجدير بالذكر أن التعرف على أخطاء متعلمي العربية كلغة ثانية في الإملاء، والنحو والدلالة له شواهد من النجاح، ولو على مستوى بعض ما أنجِزَ من الأبحاث.

٤ , ١٠- القِياس الأسلُوبيّ (Stylometry)

ومن التطبيقات والتقنيات اللطيفة إمكانية التأكد من نسبة مقال أو كتاب لكاتب معين له كتب أو مقالات أخرى، أو ما يُطلَق عليه (القياس الأسلوبيّ). والتقنية لا تعدو أن تكون فرعا من تصنيف الوثائق الذي أشرنا إليه آنفا. ويمكن الإفادة من هذه التقنية في القضايا التي تحتاج إلى توثيق نسبة مقال لكاتبه كما يحدث في نسبة الأولاد لآبائهم. فسبحان الله الذي خلقنا متشابهين ولسنا متطابقين حتى يحدث التمايز بيننا والتعارف، ليس ذلك في الجينات والبصمات فقط؛ ولكن أيضا في أسلوب الكتابة. إن الكلمات التي

يستخدمها كل منا وتصاريفها وتصاحباتها مع كلمات أخرى، هي للكاتب كالبصمة للبنان.



الشَّكل ١-٥: القِياس الأسلُوبيّ.

٤, ١١- التعرف الآليّ على الكلام (Automatic Speech Recognition)

الكلام [1] هو أفضل طرق التواصل بين البشر. وقد اكتسبت تقنية التعرف على الكلام أهميتها من ذلك. وتعد هذه التقنية خطوة نحو التواصل السهل والمريح بين الإنسان والآلة عند نضجها عمليا.

وتطبيقات هذه التقنية كثيرة ومتنوعة، منها:

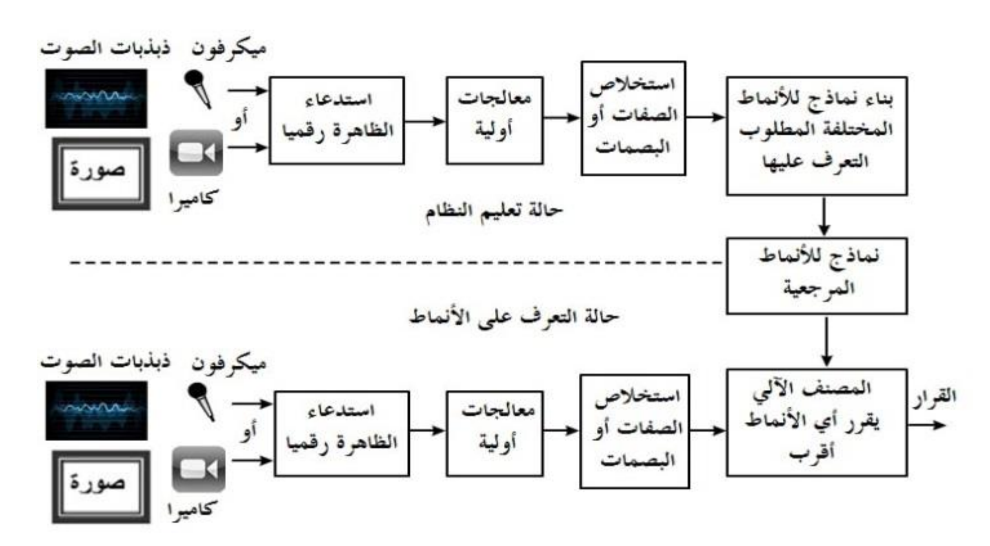
- أنظمة الإملاء (Dictation Engine)، وتستخدم لإملاء مقال أو رسالة وتحويلها من صوت إلى نص.
- تستخدم كأداة للإبحار والتجول (Navigation Tool) في بعض التطبيقات. وقد يكون ذلك مهما في مواقف مُعَيَّنة، مثل إمكانية إدارة مكالمة هاتفية بالكامل داخل سيارة وبدون الحاجة إلى البحث عن رقم الهاتف بالعين واليد، لأن ذلك يشغل عن قيادة السيارة. فتمكن هذه التقنية المستَخدِمَ من إصدار الأمر بالبحث

١- يُشير (الكلام) إلى الكلام المنطوق، ويُشير (النَّصّ) إلى النَّصّ المكتوب.

عن رقم هاتف مُعَيَّن فتقوم بهذه المهمة. كما تمكن هذه التقنية من الوصول لوثيقة أو كتاب وطلب قراءته، وعندئذ تستدعي تقنية أخرى لقراءة النصوص.

- تسجيل الاجتماعات وتحويلها آليا إلى محاضر اجتماعات: (Spontaneous Speech Recognition)، وإن كانت نتائج هذا التَّطبيق لا تزال دون المستوى المطلوب كثيرا.

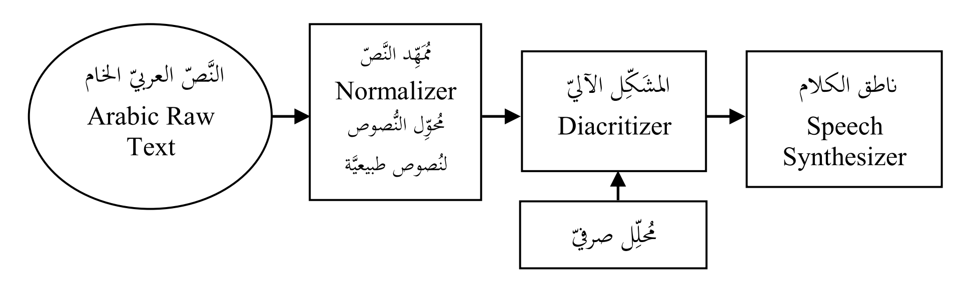
ومع نضج المجتمع الرقمي تتعدد مجالات الإفادة من تقنيات الصوت وغيره. ولكن كيف تعمل هذه التقنية؟ ببساطة فإنَّ كل تقنيات التعرف على الصوت والحروف والمتحدث تعمل تقريبا بمبادئ متقاربة. ويوضِّح الشكل (١-٦) كيفية عمل هذه الأنظمة.



الشَّكل ١-٦: نموذج عامّ للتَّعرُّف على الأنماط.

٤, ١٢ – تحويل النَّصّ إلى كلام (Text to Speech – TTS)

هذه التقنية لها تطبيقات عملية كثيرة، منها نطق الكتب للمكفوفين وضعاف البصر، ومنها نطق الرسائل عبر الهاتف للإخبار عن خدمة أو إعطاء معلومة... إلخ. وتتكون هذه التقنية من تقنيات أخرى، والشكل (١-٧) يبين مكونات تقنية تحويل النص إلى كلام العربي.



الشَّكل ١-٧: مُكَوِّنات تقنية تحويل النص إلى كلام.

يلاحظ هنا أن العربية تحتاج لمشكل آلي، وهذا الجزء غير مطلوب في اللغات اللاتينية حيث تكتب كل الأصوات غالبا. فقط تحتاج إلى بعض القواعد للنطق مثلما يحدث عند نطق كلمة آخرها (tion...).

وبالنسبة لتقنية ناطق الكلام (Speech Synthesizer) لها مدرستان، مدرسة تعتمد على تقطيع الفونيمات من كمية من الصوت المعد للتدريب إذ يُقطّع الصوت المسجل إلى فونيمات. وتُستدعي الفونيمات المطلوبة في سياقات متشابهة وقت تعليم التقنية سلفا، وهذه الطريقة هي الأكثر شيوعا.

ولكن هناك مدرسة أخرى تعتمد على توليد نماذج رياضية لكل صوت. ويمكن تدريب وتعليم هذه النماذج سلفا. المدرسة الأولى يكون الصوت فيها طبيعيا أكثر ولكن يعاني بعض التقطيع في السلاسة والانسيابية؛ والصوت الناتج عن المدرسة الثانية أكثر انسيابية ولا تشعر الأذن فيه بتقطع أجزائه، لكنَّه لا يبدو طبيعيًّا كالصَّوت النَّاتج عن المدرسة الأولى. وكلا المدرستَين تسعى إلى تحسين الأداء بزيادة كمية الصوت المستخدم في التدريب.

ولازالت الأبحاث قائمة لزيادة درجة التعبير في الصوت (Expressive Speech) حتى يتسنى للمستمع تمييز الاداء المتسم بالفرح أو الحزن أو الفزع أو الغضب ... إلخ. كما يمكن تمييز الاسلوب الخبرى من الاستفهامى ... إلخ.

٤ , ١٣ – البحث في الصوتيَّات والمرئِيَّات (Audio and Video Search)

أي البحث عن كلمة أو تعبير في ملفات صوتيَّة أو ملفات مرئِيَّة (في مسار الصوت فيها). هب أننا نريد أن نبحث في ملفات صوتية كثيرة (بالآلاف) عن «صلاة العيد» حتى نعرف الآراء المختلفة حول فقه صلاة العيد مثلا، حيث تغنينا هذه التقنية عن الاستماع إلى هذا الكم الهائل بآذاننا لنصل إلى ما نريد، حيث تقوم التقنية باستعراض كل هذا الكم من الصوت وتختار المواضع التي تشتبه في تعبير «صلاة العيد» وتدلنا عليها. ولهذه التقنية استخدامات كثيرة منها:

- البحث في الخطب والتَّسجيلات الصوتية عن موضوعات تهمنا، ويمكن أن يتم ذلك أيضا في المرئِيَّات (في مسار الصوت فيها) لاستخدامات بحثية أو ثقافية.
- محطات الإذاعة والتلفزيون تحتاجه لإعداد تقارير عن شخصية أو حدث أو مؤسسة فتستدعي ما سبق قوله حولها في تراثها أولا حتى تكون منسجمة مع نفسها فيما تعده.
- لأسباب أمنية يتم تسجيل المكالمات الدولية في كثير من دول العالم؛ والاستماع إليها يكون شاقا جدا، لذلك تُستدعي هذه التقنية وتزود بمجموعة من الكلمات التي لو وجد بعضها تنبه المعنيين بالأمر للمراجعة الدقيقة (مثال للكلمات: مخدرات، أفيون، هيروين، أسلحة،...).
- ونظرًا لتزايد حجم الصوت والمرئِيَّات على الإنترنت حاليا ومستقبلا، فإن هذه التقنية ستفيد الباحثين عن المحاضرات والخطب، بل والأفلام، وكل المطلوب منك هو كتابة عدة كلمات تذكرها من أحاديث من فيها مثلا، فتستدعي أقرب هذه الخطب أو الأفلام.

وتعتمد هذه التقنية أساسا على تقنية التعرف على الكلام، إما بصورة كاملة أو بصورة جزئية.

وجدير بالذكر أن كلمات البحث يمكن كذلك أن تكون بالصوت بدل الكتابة، وهذا يخدم مستعملي الهواتف النَّقَّالة. وكذلك تشمل هذه التقنية البحث في الصور فيكفى أن تكتب كلمة «هرم» لتستدعى آلاف الصور التى عولجت من قبل ويُظن أن بهـــا صورة «هرم» .

٤ , ١٤ – التعرف على اللُّغة (Language Recognition)

تزايدت الحاجة إلى هذه التِّقنية بعد ظهور الإنترنت والانفتاح على كل الثقافات واللغات؛ فإذا حصلنا (عند البحث عن موضوع يهمنا) على حديث صوتي له صلة بموضوع البحث، وأردنا معرفة لغة المتحدث لنعرف من أي البلاد هو كخطوة أولى نحو معرفة ما يقول. إذ نستدعي تقنية التعرف على الكلام المدربة على لغته التي يتحدث بها.

هذه التقنية يمكن شرح فكرتها باستدعاء النموذج المشروح في الشكل (١-٥) وبدلا من أن تكون الأنماط عبارة عن كلمات، تكون الأنماط عبارة عن صوت بلغات شتى، ويدرب النظام على اللغات التي يراد التعرف عليها. وبعد تكوين النماذج الرياضية لأصوات كل لغة يصبح النظام قادرًا على التمييز بينها عندما يتعرض لإحداها.

ومن الجدير بالذكر أن التعرف على اللغة من النصوص المتاحة أهون بكثير اذ يكفى أن نحتفظ بالشائع من كل لغة من الألفاظ والتعبيرات لاستخراج لغة المقال الذى نريد معرفة لغته.

٤ , ١٥ – تعلُّم النُّطق بمُساعدة الحاسوب
(Computer-Aided Pronunciation Learning)

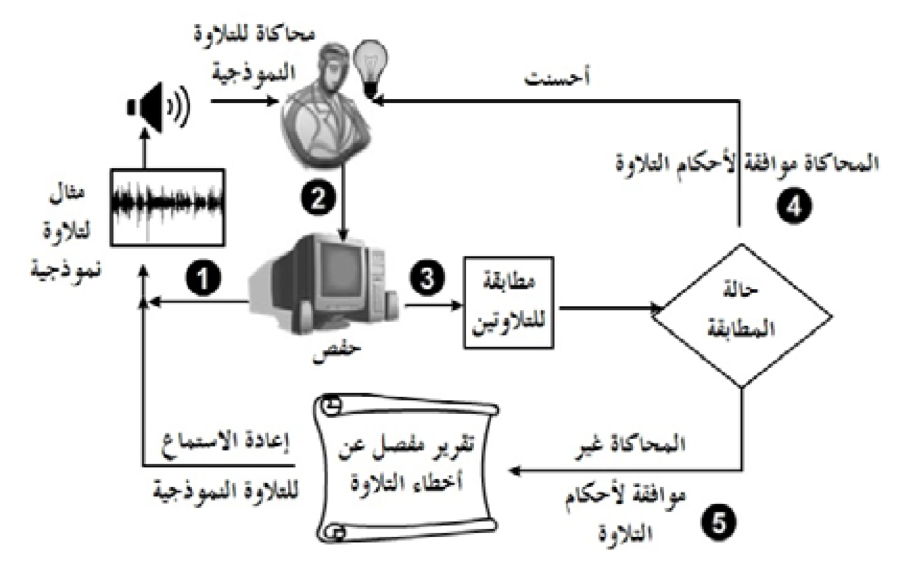
تُعَدُّ مهارة النُّطق أقل المهارات الَّتي نجيدها عبر عملية تعلم اللغات الأجنبية بالطرق التقليدية. فكم منا يجد صعوبة عندما يحادث أجنبيًّا لأول مرة أو يستمع إلى نشرة أخبار أجنبية؟ لذلك فإن لهذه التقنية مستقبل مهم في تيسير تعلم اللغات.

ويمكن أن نصنف هذه التقنية كتقنية متشابهة وغير متطابقة مع تقنية التعرف على الكلام. حيث تتفق هاتان التقنيتان في تحليل الصوت واستخلاص خصائصه وبصماته، ففي تقنية التعرف على الكلام مطلوب التعرف على الكلمة أو الجملة المنطوقة، أمَّا في هذه التقنية فالكلمة أو الجملة محل البحث معروفة سلفا، ولكن المطلوب هو التأكد من نطقها نطقا سليما. فعندئذ توضع البدائل المختلفة للنطق، الصحيح منها والخطأ، وتقوم التقنية بالتعرف على النطق.. هل كان أقرب للنطق الصحيح أو للنطق الخطأ. خذ مثالا: لو أن أجنبيا يريد تعلم العربية وليس عنده حرف مثل «الطاء» ويريد أن

ينطق كلمة «طبيب» فإننا سلفًا ندرب النظام على كلمتين «طبيب» و «تبيب» فإذا كان النطق أقرب لـ «تبيب» نعطي تغذية مرجعية للمتعلم بموضع الخطأ ونطلب منه معاودة المحاولة وربما نحيله إلى طريقة نطق «الطاء» تفصيليًّا. وهكذا يتعلم الأجنبي كيف ينطق الحروف والكلمات العربية.

وأهمية هذه التقنية تكمن في أن التدريب عليها يحتاج إلى وقت وصبر طويل قد لا يتوافر عند البشر، ولكن الحاسوب ليست لديه مشكلة في ذلك، ولا يمل التكرار الذي هو أساس في هذا الموضع.

ومن أبرز التطبيقات التي ظهرت في هذا المجال تقنية لتعليم أحكام التجويد تحت اسم «حفص» وتعد خير مثال كتطبيق نافع في مجال مهم.



الشَّكل ١-٨: تقنية «حفص» - تعلُّم النُّطق بمُساعدة الحاسوب.

٤,١٦ - التعرف الآليّ على الكتابة المطبوعة (Typewritten OCR)

تتجه البشرية إلى إتاحة تراثها في شكل رقمي حتى يسهل التعامل معه من استدعاء وتلخيص ونطق آلي....إلخ. لذلك نشأت أهمية هذه التقنية. إن كمية الوثائق الحكومية وغير الحكومية وكمية الرسائل العلمية التي تعمر مكتبات مئات الجامعات في المنطقة العربية تحتاج إلى تقنية عالية الجودة لتيسير رقمنة هذه الوثائق.

وتقوم فكرة تقنية التعرف الآليّ على الكتابة المطبوعة (مع اختلاف الخطوط نوعا وحجما) بمسح ورقات الوثيقة أو الكتاب آليًّا فيصبح صورة رقمية، ثم تقوم التقنية باستخراج الأجزاء من الصور، والتي يوجد بها نصوص مطبوعة ثم تفصل الأسطر ثم الكلمات (وربما الحروف)، ويتمُّ التَّعرُّف على (الحروف ثُمَّ) الكلمات فالجمل فالفقرات، والشكل (١-٩) ليبين الفكرة الأساسية.



الشَّكل ١-٩: التَّعرُّف الآليّ على الكتابة المطبوعة.

٤ ,١٧- التعرُّف الآليّ على الكتابة المخطوطة (OCR Handwriting)

وتُشبه هذه التقنيةُ سالفتَها، إلا أن الخطوط اليدوية لها تحدياتها ومُشكلاتها المختلفة نوعا ما عن مُشكلات الكتابة المطبوعة. المشكلة المراد حلها هنا هي نفس المشكلة المراد حلها في تقنية التعرف على الكتابة المطبوعة، ولكن المادة المستخدمة للتدريب والتعرف مختلفة كثيرا. ففي حالتنا قد لا تكون الأسطر والحروف منتظمة، حتى لنفس الكاتب وفي نفس الوثيقة، حيث يختلف شكل الحرف في كل مرة يكتب فيها (على خلاف الكتابة المطبوعة).

٤ ,١٨- التَّعَرُّف الآنيّ على الحُرُوف (Intelligent Character Recognition)

بعد انتشار الهواتف النَّقَّالة ذات الشاشات الحساسة للَّمس، ومن قبل ذلك ظهور الحواسيب اللَّوحِيَّة [التي يمكن الكتابة عليها] (Tablet computes)، ظهرت الحاجة لمثل هذه التقنية لأنَّ الكتابة بالقلم عند عامة الناس أسهل من استخدام لوحة المفاتيح. وإذا كان استخدام لوحة المفاتيح في حالة الحاسوب ممكنًا لاتساع المكان وطول الدربة،

فإن الكتابة باستخدام لوحة المفاتيح على الهواتف النَّقَّالة أصعب ولذلك تكتسب هذه التقنية أهمية أكبر في هذه الحالة.

تتميز ظروف هذه التقنية بمواصفات مواتية، منها أنه لا توجد ضوضاء كما في الصفحات المصورة سواء لكتابة اليد أو الكتابة المطبوعة. كما تتوفَّر لدينا معلومات أكثر عن وضع القلم وترتيب الكتابة قد يكون مفيدا عند التطبيق. وتتفق هذه التقنية مع سالفتيها في نفس النموذج المرجعي في الشَّكل (١-٩) الوارد آنفا.

٤,١٩ - تعلُّم الخَطّ بمُساعدة الحاسوب
(Computer-Aided Font Learning)

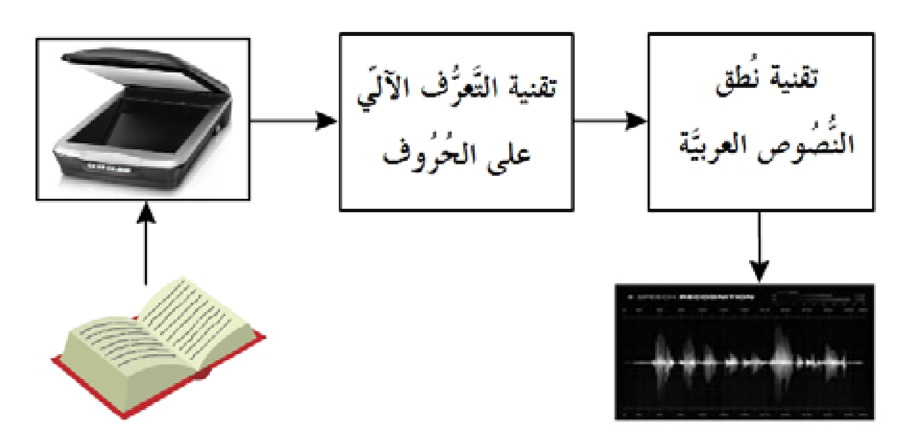
هذه التقنية تشبه تقنية تعليم النطق السليم للغة أجنبية. فهي تساعد على وزن نسب الأطوال والزوايا المختلفة لكتابة الحروف وتراكيبها للغة ما. إن توظيف الحاسوب للقيام بهذه المهمة يعد استخدامًا مناسبًا لرفع كفاءة المدارس في تحسين خطوط الطلاب. ويمكن بالطبع الانتفاع بها لكل الأعمار ولمن سيتعلم اللغة من الأجانب أو العكس عند تعلم العربيّ لكتابة الحروف اللاتينية أو الصينية مثلا.

وكما قيل في تعليم النطق يقال هنا. إن أصل ومرجع التقنية هو الشكل (١-٧)، مع وضع مواصفات لمقاييس الحرف المقبول؛ فإذا تم الالتزام بها يعطي الكاتب تغذية مرجعية إيجابية؛ أما إذا خالف في تلك المقاييس فيتم التعرف علي موضع الخلل ويعطي تغذية مرجعية عن الخلل وموضعه. ولا يملّ الحاسوب (أو الهاتف النَّقَّال) من طول التدريب حتى تُصقل المهارة ويتحسن الخط.

تقنيات مركبة

٤,٢٠ - تقنية قراءة الكتب (Book Reader)

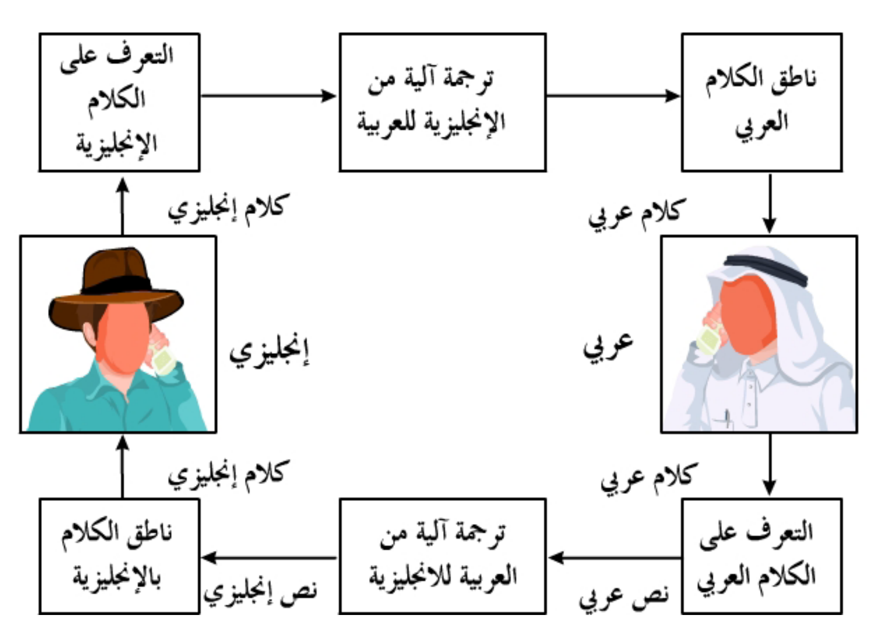
وتتركب هذه التقنية من تقنيتين فَرعِيَّتَين، حيثُ تُعنى الأولى بالتَّعَرُّف على الحُرُوف بعدَ إدخالِها إلى الحاسُوب في هيئة صُوَرٍ باستخدام الماسِحات الضَّوئِيَّة، ليتَحَوَّلَ شكلُ المستَنَد - بذلكَ - من صُورةٍ ورقيَّةٍ إلى صُورةٍ رقميَّة يُمكنُ التَّحكُّمُ فيها؛ وتُعنى التِّقنيةُ الأخرى بقراءة النُّصُوص من خلال تحويل النَّصّ (المكتوب) إلى كلامٍ (منطوق).



الشَّكل ١-١٠: تقنية قراءة الكتب.

٤, ٢١- تقنية التَّرجمة الشَّفَهِيَّة الآليَّة (Speech to Speech Translation)

وهي تقنية معقدة إلى درجةٍ كبيرة، إذ تتكون من ٦ تقنيات على النَّحو المبَيَّن بالشكل (١-١١):

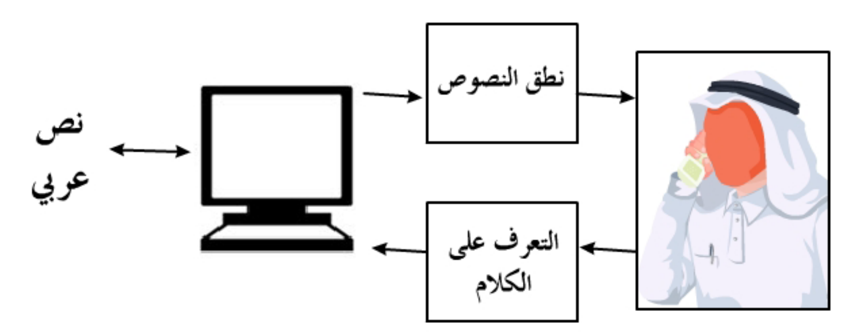


الشكل ١-١١: تقنية التَّرجمة الشَّفَهِيَّة الآليَّة.

٤ , ٢٢- أنظمة الاستجابة الصَّوتِيَّة التَّفاعُلِيَّة

(Interactive Voice Response System – IVR)

كما في الشَّكل (١-١٢)، فإن تقنيتي التعرف على الكلام ونطق النصوص تستخدمان للتفاعل مع المستخدم لحجز تذكرة أو معرفة معلومة أو أداء خدمة.

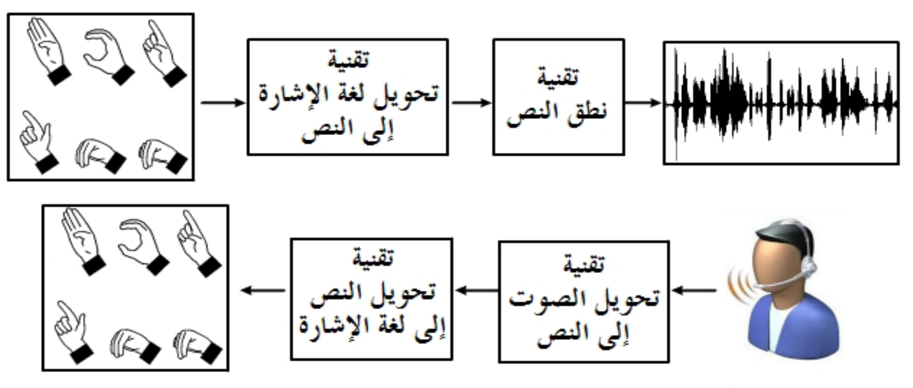


الشَّكل ١-١٢: أنظمة الاستجابة الصَّوتِيَّة التَّفاعُلِيَّة.

٤ , ٢٣- تقنية تحويل الكلام إلى لُغة الإشارة، والعكس

(Speech-to-Sign-Language & Sign-Language-to-Speech)

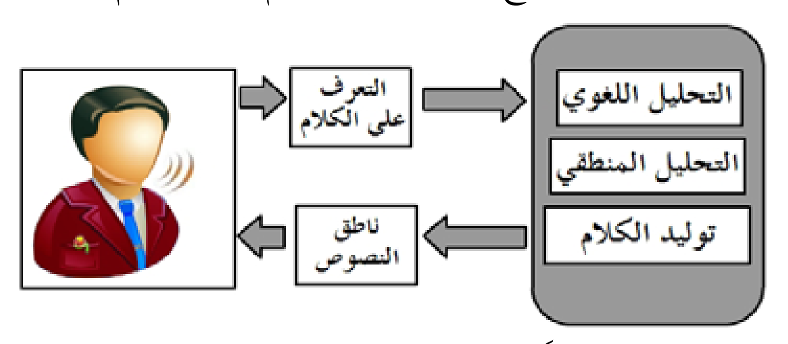
في هذه الحالة تستخدم قفازات خاصة بها حساسات ترسل إشارات للحاسوب بموضع كل سُلامي في أصابع اليدين. وباستخدام تقنية تحويل إشارات القفازات إلى نص، يمكن استخدام تقنية نطق النصوص لتحويلها إلى كلام والعكس، إذ يمكن تحويل الكلام إلى إشارات للصم والبكم. فالمتحدث يحول كلامه إلى نص باستخدام تقنية التعرف على الكلام، ثم تحول إلى إشارات على شاشة الحاسوب، وهكذا يمكن وصل الصم والبكم بالمجتمع، فقط نحتاج إلى قفازات بها حساسات وميكروفونات وحاسوب أو نقَّال يمكن أن تُحمَّلَ عليه التقنيات؛ كما أن هناك جهود بحثية للتعامل مع صورة اليدين عبر كاميرا ولكن لا تزال نتائجها دون المستوى التطبيقيّ.



الشَّكل ١-١٣: تقنية تحويل لُغة الإشارة إلى كلام، والعكس.

٤ , ٢٤- أنظمة إدارة الحوار (Dialog management Systems)

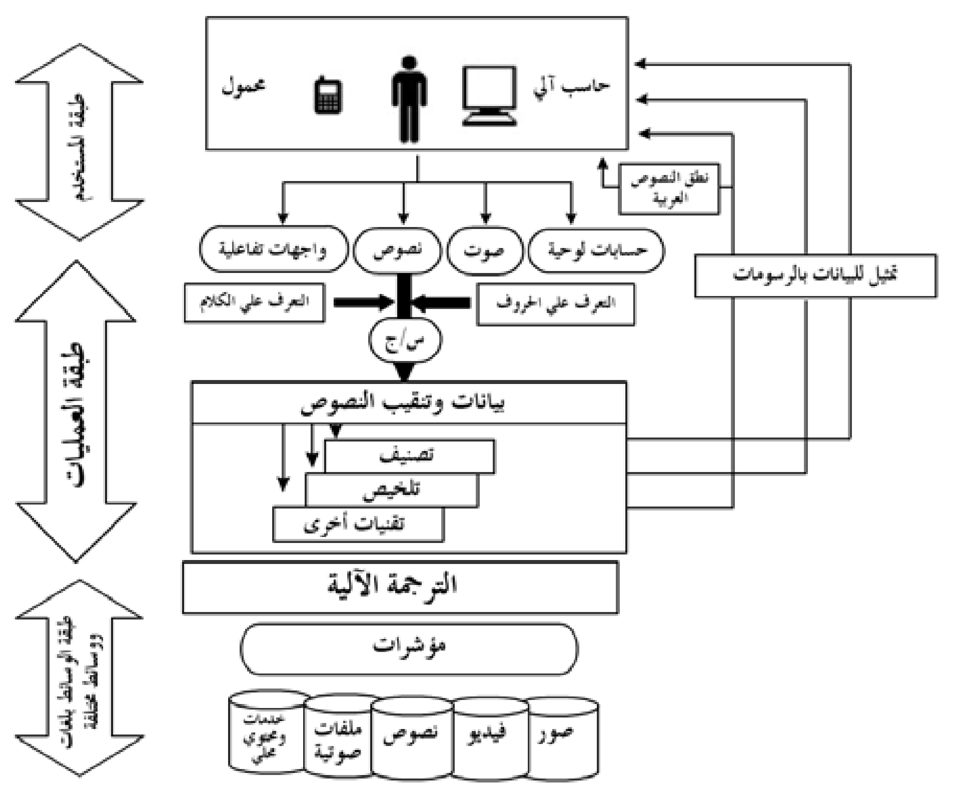
وتشبه أنظمة التفاعل الصوتي، والتي تحتاج إلى تقنيات التعرف على الكلام ونطق النصوص، ولكن يضاف إليها تقنية عميقة لفهم الكلام وتقنية أخرى لتوليد الكلام. في الواقع هناك طبقة مبسطة ومبرمجة لأنظمة التفاعل الصوتي IVR، لفهم وتوليد الكلام ولكن إدارة الحوارات تحتاج لدرجة أعمق لفهم الكلام وتوليد الإجابات. نحتاج إلى محللات صرفية ونحوية ودلالية «ومحلل منطقي» لما يقال. هذه الأنظمة ستنقل الآلة لمستويات أعلى بكثير من حيث سهولة التفاعل مع الإنسان وحسن توظيف إمكانياتها لخدمته. وسوف تستخدم بكثافة مع «الإنسان الآلي» (Robot) في ثوبه الحديث، إذ سيكون في مقدور هذه الآلة الذكية «الإنسان الآلي» القيام بكثير من الأعمال الروتينية في المنزل والمصنع وخدمة الأطفال وكبار السن والمرضى على مدار ٢٤ ساعة في اليوم دون كلل ولا ملل. بل ربما يكون قادرا على الحكاية والترفيه لإخراج مستخدميه من همومهم واللعب معهم بذكاء وحضور.



الشَّكل ١-١٤: أنظمة إدارة الحوار.

٤,٢٥- الأنظمة المتَقَدِّمة لاسترجاع المعلومات
(Advanced Information Retrieval Systems)

لا تكتفي هذه الأنظمة باسترجاع المعلومات المخَزَّنة في أوعية المعلومات في صُورَتِها المباشِرة؛ وإنَّما يمتدُّ عملُها ليشملَ - كذلكَ- استرجاعَ معلوماتٍ مُعَقَّدة من مجموعاتٍ ضخمةٍ من الوثائِق والأوعِية؛ بالإضافة إلى قُدرة هذه الأنظمة على تلخيص المعلومات المستَرجَعَة إذا دَعَت الحاجةُ إلى ذلك، أو استرجاعها من لُغاتٍ مُختلفة، أو استخدام أساليب مُتَقَدِّمة في عمليَّات الاسترجاع، كأن يُستَخدَمَ الصَّوتُ عبرَ الهاتِف النَّقَّال أو اللَّمس في الحواسيب اللَّوحِيَّة، ونحو ذلك كما في الشَّكل (١-١٥).



الشَّكل ١-١٥: الأنظمة المتَقَدّمة لاسترجاع المعلومات.

٤, ٢٦- التَّنقيب في النصوص (Text Mining)

تهدفُ هذه التِّقنية إلى استخلاص معلوماتٍ عالية الجودة من البيانات النَّصِّيَّة. وتتمُّ عبرَ إجراءاتٍ، تبدأ بتحليل النُّصُوص المدخَلَة وإعدادِها بإضافة معلوماتٍ لُغَوِيَّةٍ مُصاحِبة، ثُمَّ إدراجها في قواعد بياناتٍ نَصِّيَّة؛ ومرورًا باستخلاص المعلومات من قواعد البيانات؛ وانتهاءً بتقييم النَّتائِج. وتتفرَّعُ عن هذه التِّقنية تقنياتٌ أخرى، منها:

- تحليل المشاعِر.
- تلخيص وتصنيف وتجميع النُّصُوص.
- التَّنقيب عن اتِّجاهات الرَّأي.
- التصحيح الآلى للنصوص.

٥- من المؤسَّسات المعنِيَّة بحوسبة اللُّغة العربيَّة

ثمَّةَ العديدُ من المؤَسَّسات المعنِيَّة بحوسبة اللُّغة، والتي تعمل في خدمة تقنيات اللغة العربية، سنُحاول أن نعرِضَ لبعض هذه المؤَسَّسات فيما يلي:

٥, ١- مؤسَّسات ذات طابع أكاديميّ

- **مُؤَسَّسة LDC (Linguistic Data Consortium):**

هي مُؤَسَّسة بحثيَّة، تُساهم فيها مجموعةٌ من الجامعات والمؤَسَّسات المعنِيَّة بالموارد اللُّغَوِيَّة، بالإضافة إلى عددٍ من المختَبَرات البَحثِيَّة الحُكُومِيَّة. تأَسَسَت في عام ١٩٩٢ بدعمٍ من وكالة مشروعات البُحُوث المتَقَدِّمة (the Advanced Research Projects Agency –ARPA). تولي مُؤَسَّسة LDC عنايةً ببناء المدَوَّنات اللُّغَوِيَّة، وقواعد البيانات اللُّغَوِيَّة المكتوبة والمنطوقة، والمعاجم اللُّغَوِيَّة لأغراض البحث والتَّطوِير.

- المقرّ: جامعة بنسلفانيا – الولايات المتَّحدة الأمريكيَّة.
- الموقع الإلكترونيّ: http: / /www.ldc.upenn.edu

- مِن الموارد اللُّغَويَّة للعَرَبِيَّة:

– المحلِّل الصَّرفِيّ للُّغة العربيَّة Buckwalter (مُزَوَّد بمُعجمٍ لُغَوِيّ).

– البنك النَّحويّ العربيّ Treebank Arabic.

– مُعجم اللُّغة العربيَّة الدَّارِجة في مِصر (مُعجم العامِّيَّة المِصرِيَّة).

– العديد من المدَوَّنات اللُّغَوِيَّة للعامِيَّات العربيَّة (في مصر والعراق).

- **فريقُ البحث في مُعالجَة اللُّغات الطَّبِيعِيَّة، كُلِّيَّة عُلُوم الحاسوب، ليدز:**

(NLP - School of Computing at the University of Leeds)

تُعَدُّ كُلِّيَّة عُلُوم الحاسوب في جامعة ليدز مُؤَسَّسَةً أكادِيميَّةً مُتَخَصِّصة، تمنحُ الدَّرجات العلمِيَّة للباحِثِين، وتُشرِف على أطروحاتِهم للماجستير والدُّكتوراه. والكُلِّيَّةُ مُصَنَّفةٌ ضمن أفضل ١٠ مؤسَّسات أكادِيمِيَّة مُتَخَصِّصة في علوم الحاسوب بالمملكة المتَّحدة. يعمل فريقُ البحث في مُعالجَة اللُّغات الطَّبِيعِيَّة بالكُلِّيَّة في مشروعاتٍ بحثِيَّة مُشتَرَكة، ويُولي عنايةً خاصَّةً بتقنيات اللُّغة العَرَبِيَّة وموارِدِها اللُّغَوِيَّة. يُرَكِّزُ الفريقُ في أبحاثِه على نمذَجة اللُّغة وبناء المدَوَّنات اللُّغَوِيَّة العربِيَّة.

- المقرّ: جامعة ليدز – المملكة المتَّحدة.

- الموقع الإلكترونيّ:

http: / /www.engineering.leeds.ac.uk /computing /postgraduate / research–degrees /projects /natural–language–processing.shtml

- مِن الموارد اللُّغَويَّة للعَرَبِيَّة:

– المدَوَّنة اللُّغَوِيَّة للعربيَّة المعاصِرة.

– المعَنوِن التِّلقائِيّ لأقسام الكلام في اللُّغة العربيَّة المكتوبة.

– نموذَج حاسُوبيّ للتَّمثِيل المعرِفيّ في القُرآن الكريم.

– المدَوَّنة اللُّغَوِيَّة العربيَّة لنُصُوص القُرآن الكريم.

- **المنَظَّمة العَرَبيَّة للتَّربية والثَّقافة والعُلُوم (ألكسو)- إدارة العُلُوم:**

The Arab League Educational, Cultural and Scientific Organization – ALECSO

مع مطلع الألفيَّة الثَّالثة وجَّهَت إدارةُ العُلُوم بالمنَظَّمة العَرَبيَّة للتَّربية والثَّقافة والعُلُوم عنايتَها إلى حوسَبة اللُّغة العربيَّة، فعقدَت عددًا من المؤتمرات والمحافل الدَّوليَّة، وأنجزَت – ولا تزال – عدَّةَ مشرُوعاتٍ تتبعُ مجالَ المعالَجة الآليَّة للُّغة العَرَبيَّة. يعملُ في مشروعات حوسبة اللُّغة بإدارة العُلُوم عددٌ من الأكاديمِيِّين والخُبراء العَرَب من خلال فِرَق عملٍ ومجموعاتٍ بحثِيَّة، كما تتعاونُ مع المنَظَّمة مجموعة من المؤَسَّساتُ والهيئات المعنِيَّة بحوسبة اللُّغة العَربِيَّة. تجدُر الإشارةُ إلى عناية المنَظَّمة العربيَّة للتَّربية والثَّقافة والعُلُوم بتوفير تقنيات اللُّغة العربيَّة المنجَزَة في صُورةٍ حُرَّة أو مفتوحة المصدر، سعيًا إلى إتاحتها للباحثين من ناحية، والعمل على تطويرِها ومُعالَجة جوانب القُصُور فيها من ناحيةٍ أخرى.

- المقرّ: تونُس.
- الموقع الإلكترونيّ: http: / /www.alecso.org
- مِن الموارد اللُّغَوِيَّة للعَرَبِيَّة:

– نظام التَّحليل الصَّرفِيّ «الخليل».

– المعجم العربيّ التَّفاعُلِيّ.

- مِن المشروعات القائمة:

– نظام التَّحليل النَّحوِيّ.

– المدَقِّق الإملائيّ للُّغة العربيَّة.

– آليَّة تشكيل النُّصُوص العَرَبِيَّة.

- **مدينة الملك عبد العزيز للعُلُوم والتِّقنية (معهد بُحوث الحاسب):**

(King Abdulaziz City for Science and Technology)

تأسَّسَ المعهدُ في عام ١٩٩٢ لأغراضِ البحثِ الَّتي تخدم المملكةَ العربيَّة السُّعوديَّة وخُطَطَها للعُلُوم والتِّقنية في مجال تقنية المعلومات. ويَضُمُّ المعهدُ أربعةَ أقسام علميَّة، هي: قسم الصَّوتيَّات واللُّغَوِيَّات، وقسم الحوسبة العلمية، وقسم النظم والشَّبكات، وقسم هندسة البرمجيات والأنظمة المبتكرة. يُعنى قسم الصَّوتِيَّات واللّغَوِيَّات بإعداد البُحوث والدِّراسات وتقديم الحُلول لُمشكلات حوسبة اللُّغة العربيَّة، كما يقوم بتقديم الاستشارات وإقامة المؤتمرات واللِّقاءات العلميَّة المَتَخَصِّصة لُمتابعة التَّطوُّرات الحادثة في ميادين تقنيات اللُّغة والبحث في سُبُل التَّعاوُن مع مراكز الأبحاث والتَّطوير الدولية.

- المقرّ: مدينة الملك عبد العزيز للعُلُوم والتِّقنية - الرِّياض - المملكة العربيَّة السُّعُودِيَّة.

- الموقع الإلكترونيّ:

http://www.kacst.edu.sa/ar/about/institutes/pages/ce.aspx

- مِن الموارد اللُّغَوِيَّة للعَرَبِيَّة:

- البنك السُّعُودِيّ للأصوات.
- المقَوِّم الآلِيّ للمقالات العربيَّة.
- نظام مُعالجة الحُرُوف العربيَّة المطبوعة.
- مجموعة نظم التَّحليل النَّحوي للنُّصوص العربيَّة المتاحة على الإنترنت.
- مدونة عربية نصية كبيرة.

- مِن المشروعات القائمة:

- مشروع بناء نظام تفاعلي للتَّعرُّف على الأصوات العربيَّة.
- مشروع ويكي عربي - لإثراء موسوعة ويكيبيديا العربية.

- **مجموعة كولومبيا لنمذَجة اللَّهجات العَرَبِيَّة، جامعة كولومبيا:**

(«Columbia's Arabic Dialect Modeling Group» CADIM)

مجموعةٌ بحثِيَّةٌ كوَّنَها فريقُ البحث في مركز أنظمة التَّعلُّم الحاسُوبيَّة (Center for Computational Learning Systems –CCLS) بجامعة كولومبيا الأمريكِيَّة. يُعنى الفريقُ بمُعالَجة اللَّهجات العربِيَّة، ويستندُ إلى معايير اللُّغة العرَبيَّة المعاصِرة (Modern Standard Arabic –MSA). تَبَنَّى المجموعةُ مشروعًا للتَّعرُّف الآليّ على الكلام (Arabic Automatic Speech Recognition –AASR)، ويهدف المشروع – عُمومًا – إلى تعيين أوجه التَّبايُن بين اللَّهجات العربِيَّة على مُستوى الأصوات والتَّراكِيب والمعجم. كما امتد اهتمام المجموعة للترجمة الآلية من العربية إلى الإنجليزية والتعامل مع اللهجات العربية.

- المقرّ: جامعة كولومبيا – نيويورك – الولايات المتَّحِدة.
- الموقع الإلكترونيّ: http://www1.ccls.columbia.edu/~cadim/
- مِن الموارد اللُّغَوِيَّة للعَرَبِيَّة:

- نظام التَّحليل اللُّغَوِيّ MADA+TOKAN. وهُو نِظامٌ لتحليل النُّصُوص العربِيَّة المكتوبة، ومن مهامِّه:
- تَقطِيع النُّصُوص (Tokenization).
- التَّشكِيل الآليّ (Diacritization).
- فَكّ الالتباس الصَّرفِيّ (Morphological Disambiguation).
- عمليَّات تَعيين أقسام الكلام (PoS Tagging)، والتَّجذِيع (Stemming)، والتَّفرِيع (Lemmatization).

- **فريقُ البحثِ في مُعالجةِ اللُّغَاتِ الطَّبيعيةِ بِجَامِعَة ستَانْفُورْد:**

(The Stanford NLP Group)

جَامِعة ستَانْفُورْد هي جَامِعة أَمْرِيكية خَاصَّة تَم افتتاحُها في ١ أُكْتُوبر ١٨٩١. يَتَألَّف فَرِيق ستَانْفُورْد لِمُعَالجة اللُّغاتِ الطبيعيةِ من قِسْمَيْ اللسانيات وعُلوم الحَاسِبِ، والذي

يَعْمَل أَعْضَاؤه معًا على الخُوارِزْميات التي تَسْمَح لِأجهزة الكمبيوتر بمعالجَةِ وَفَهْم اللغات البشرية.

- المقرّ: كَالِيفُورنيا – الولايات المتحدة الأمريكية.
- الموقع الإلكترونيّ: http: / /nlp.stanford.edu
- من المشروعات:
  - نِظَام ستَانْفُورْد للتَّرْجَمة الآلية.
  - نِظَام ستَانْفُورْد لاستنتاجِ مُدَوَّنات اللُّغات الطبيعية.
  - نظام ستَانْفُورْد المُفْتُوح لِإسْتِخْراج البَيَانات.

▪ **مَعْهد قَطَر لِبُحوث الحُوْسَبة (QCRI)**

(Qatar Computing Research Institute)

هُو مَعْهد بَحْثِيٌّ غَير رِبْحيٍّ، تَأَسس كَجُزْءٍ من مؤسسة قَطَر للتربية والعلوم وتنمية المجتمع في عام ٢٠١٠، ومن تَخَصُّصَاته التي يُركز عليها تِقْنِيَّات اللغة العربية والحوسبة عَالية الأداء والمعلوماتية الحيوية.

- المقرّ: الدَّوْحَة – قَطَر.
- الموقع الإلكترونيّ: http: / /qcri.org.qa
- تَحدِّيات المُعْهد في مَجَال تِقْنِيَّات اللغة العربية:
  - تَحدِّي التَّرْجَمة الآلية للغة العربية.
  - تَحَدِّي مُعَلِّم اللغة العربية الإلكترونيّ.
  - تَحَدِّي نِظَام المُحَادَثَة المُستمرَّة باللغة العربية.
  - تَحَدِّي المحتوى وَالبَحْث باللغة العربية.

▪ **الجمعِيَّة المِصريَّة لهندسة اللُّغة** ESOLE:

(The Egyptian Society of Language Engineering)

أُنشِئَت في عام ١٩٩٦ بكُلِّيَّة الهندسة في جامعة عَين شمس، وتُعنى بحوسبة اللُّغَة عموما واللُّغة العربيَّة على وجه الخُصُوص. تعملُ الجمعِيَّة – وفقًا لأهدافها المعلَنة – على تحقيق التَّقارب بين الأكاديميِّين والباحثين والتِّقنيِّين والمنتجين والمستخدمين من مختلف الأماكن ومختلف اللُّغات ممَّا يساعد على تبادل المعلومات من أجل تحسين تقنية الاتِّصال بين الإنسان والآلة باللُّغة الطَّبيعيَّة. وتتمثَّلُ جوانب العناية بتقنيات اللُّغة العربيَّة في مُتابعة التَّطَوُّر التِّقَنِيّ في ميادِين مُعالجَتِها آليًّا، وفتح قنواتٍ للتَّعاوُن البحثِيّ والأكادِيميّ مع الجمعيَّات العلميَّة المعنِيَّة بحوسبة اللُّغة على المستوى المحلِّيّ والقومِيّ والدَّوليّ.

- المقرّ: جامعة عَين شمس – القاهرة – مِصر.
- الموقع الإلكترونيّ: http: / /www.esole-eg.org
- مِن أنشطة الجمعِيَّة:

– عقد مؤتمرٍ سنويٍّ لهندسة اللُّغة، يُناقِشُ مختلف القضايا في حَوسبة اللُّغة وتطبيقاتها.

– عقد ندواتٍ وورش عمل للإفادة من خبرات المتَخَصِّصِين في مجالات حوسبة اللُّغة. وتنظيم دوراتٍ تدريبيَّة للمَعنِيِّين بحوسبة اللُّغة.

▪ **مركز تِقنيات اللُّغة العربيَّة** ALTEC:

(Arabic Language Technology Center)

مُؤَسَّسَة أهلِيَّة لا تهدِف إلى الرِّبح. تأَسَّسَ المركز في عام ٢٠٠٩ بدعمٍ من شركاتٍ تِقَنِيَّةٍ ومُؤَسَّساتٍ بحثِيَّةٍ وأكادِيميَّة، ويهدف إلى توفير الموارِد اللُّغَوِيَّة للمَعنِيِّين بحوسبة اللُّغة العربيَّة من الباحِثين والمطَوِّرين، كما تهدف إلى تنمية الموارِد البشريَّة من خلال التَّدريب على مجالات هندسة اللُّغة العربيَّة والسَّعي إلى إيجاد درجاتٍ علميَّةٍ مُتَخَصِّصة بالجامعات والمؤسَّسات الأكادِيميَّة. ويُعنى المركزُ بتقديم الخدمات التِّقَنِيَّة لذوي الاحتياجات الخاصَّة من خلال تطوير التِّقنيات المناسبة لاحتياجاتِهم، كما يُعنى بتطوير التِّقنيات المساعدة في تعلُّم اللُّغة العربيَّة لغير النَّاطِقِينَ بها.

- المقرّ: الجيزة - مِصر.
- الموقع الإلكترونيّ: http: / /www.altec-center.org
- مِن أنشطة المركز:

- عقد مؤتمرات في حوسبة اللُّغة العَرَبِيَّة وتقنياتِها.
- إنجاز مجموعة من الأنظمة والموارد اللغوية للباحثين والمطَوِّرين.

- مِن الموارد اللُّغَوِيَّة للعَرَبِيَّة:

- مجموعة ضخمة من قواعد البيانات المستَخدَمة في التَّعرُّف على الحُرُوف العربيَّة - سواءٌ أكانَت ممسوحةً ضوئِيًّا (يحتوى على - ١٤٠٠٠ صفحة) أم مكتوبةً باليد (كتابه لآلف كاتب كتابة آنيّه on-line).
- مدَوَّنة لُغَوِيَّة مضبوطة بالشَّكل الكُلِّيّ، أُنجِزَت باستخدام آليَّة التَّشكيل الآليّ للنُّصُوص العَرَبِيَّة، وكذلك محشاة بأسماء الكائِنات ويصل حجم المدونة إلى ٣ملايين كلمة.

٥,٢- شركات مهتمة بحوسبة اللغة

وتجدُر الإشارةُ - كذلك - إلى وُجود العديد من الشَّركات المعنِيَّة بحوسبة اللُّغة؛ منها - على سبيل المثال لا الحَصر:

- **شركة «آي بي إم»:**

(IBM (International Business Machines Corporation

- المقرّ الرَّئيس: نيويورك - الولايات المتَّحِدة.
- الموقع الإلكترونيّ: http: / /www.ibm.com

- **شركة «مايكروسوفت»** Microsoft:

- المقرّ الرَّئيس: واشنطن - الولايات المتَّحِدة.
- الموقع الإلكترونيّ: http: / /www.microsoft.com

- **شركة «صَخر» Sakhr Software - Arabic Language Technology:**
  - المقرّ الرَّئِيس: القاهرة – مِصر.
  - الموقع الإلكترونيّ: http: / /www.sakhr.com
- **الشّركة الهندسيَّة لتطوير النُّظُم الرَّقميَّة «آر دي آي»:**
(RDI (The Engineering Company for Digital Systems Development
  - المقرّ الرَّئِيس: الجيزة – مِصر.
  - الموقع الإلكترونيّ: http: / /www.rdi–eg.com

٦- أفكارٌ بحثيَّة لأطرُوحاتٍ علميَّةٍ مُستَقبليَّة

١- إعداد دراسة عن أهمّ التِّقنيات الَّتي قُدِّمَت في حوسبة اللُّغة عمومًا وحوسبة اللُّغة العربيَّة على وجه الخُصُوص.

٢- إعداد دراسةٍ عن أهم التطبيقات لهذه التقنيات في حياة الناس.

٣- البحث في مستقبل هذه التقنيات والتطبيقات وخاصة في عصر الهواتف النَّقَّالة، وما قد يتطلبه ذلك من إعادة النظر في أوَّليات المعالجات اللغوية.

ببليوجرافيا مرجعيَّة

1. Alansary, S. & Nagi, M. & Adly, N. (2013): A Suite of Tools for Arabic Natural Language Processing: A UNL Approach. (ICCSPA'13), Sharjah, UAE.
2. Benesty, J. & Sondhi, M. M. & Huang, Y. (2007): Handbook of Speech Processing. Springer.
3. Clark, A. & Fox, C. & Lappin, S. (2010): The Handbook of Com-putational Linguistics and Natural Language Processing. John Wiley & Sons.
4. Deng, L.; Liu, Y. (2018). Deep Learning in Natural Language Processing. Springer.
5. Dickinson, M. & Brew, C. & Meurers, D. (2012): Language and Computers. John Wiley & Sons.
6. Dry, H. A. & Lawler, J. (2012): Using Computers in Linguistics: A Practical Guide. Routledge.
7. Dybkjr, L. & Hemsen, H. & Minker, W. (2007): Evaluation of Text and Speech Systems. Springer.
8. Farghaly, A. A. S. (2010): Arabic Computational Linguistics. University of Chicago Press.
9. Johnson, M. & Khudanpur, S. P. & Ostendorf, M. & Rosenfeld, R. (2004): Mathematical Foundations of Speech and Language Processing. Springer.
10. Jurafsky, D. & Martin, J. H. (2009): Speech And Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Prentice Hall.
11. Kao, A. & Poteet, S. R. (2007): Natural Language Processing and Text Mining. Springer.
12. Karwowski, W. (2019). Intelligent Human Systems Integration 2019. Springer.

13. Kulkarni, A.; Shivananda, A. (2019). Natural Language Processing Recipes: Unlocking Text Data with Machine Learning and Deep Learning using Python. Apress.

14. Kumar, E. (2011): Natural Language Processing. I. K. International Pvt Ltd.

15. Manning, C. D. & Schütze, H. (1999): Foundations of Statistical Natural Language Processing. MIT Press.

16. Manning, C. D. & Surdeanu, M. & Bauer, J. & Finkel, J. & Bethard, S. J. & McClosky, D. (2014): The Stanford CoreNLP Natural Language Processing Toolkit. Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp. 55-60, Baltimore, Maryland.

17. Mihalcea, R. & Radev, D. (2011): Graph-based Natural Language Processing and Information Retrieval. Cambridge University Press.

18. Neamat El, G.; Yee, S. (2018). Computational Linguistics, Speech and Image Processing for Arabic Language. World Scientific.

19. Ovchinnikova, E. (2012): Integration of World Knowledge for Natural Language Understanding. Atlantis Press.

20. Pasha, A. & Al-Badrashiny, M. & Diab, M. & El Kholy, A. & Eskander, R. & Habash, N. & Pooleery, M. & Rambow, O. & Roth, R. M. (2014): Madamira: A fast, comprehensive tool for morphological analysis and disambiguation of Arabic. Proceedings of the Language Resources and Evaluation Conference (LREC), Reykjavik, Iceland.

21. Pustejovsky, J. & Stubbs, A. (2012): Natural Language Annotation for Machine Learning. O'Reilly Media, Inc.

22. Rkowski, A. P. & Piasecki, M. & Przepiórkowski, A. & Jassem, K. & Fuglewicz, P. (2012): Computational Linguistics: Applications. Springer.

23. Rosner, M. & Johnson, R. (1992): Computational Linguistics and Formal Semantics. Cambridge University Press.
24. Sharp, B.; Sedes, F.; Lubaszewski, W. (2017). Cognitive Approach to Natural Language Processing. Elsevier.
25. Shoufan, A. & Al-Ameri, S. (2015): Natural Language Processing for Dialectical Arabic: A Survey. Proceedings of the Second Workshop on Arabic Natural Language Processing, pp. 36-48, Beijing, China.
26. Srinivasa-Desikan, V. (2018). Natural Language Processing and Computational Linguistics: A practical guide to text analysis with Python, Gensim, spaCy, and Keras. Packt Publishing.
27. Tavast, A. & Muischnek, K. & Koit, M. (2012): Human Language Technologies - the Baltic Perspective: Proceedings of the Fifth International Conference Baltic HLT 2012 - Frontiers in Artificial Intelligence and Applications. IOS Press.
28. Vetulani, Z. "Ed". (2011): Human Language Technology. Challenges for Computer Science and Linguistics: 4th Language and Technology Conference, LTC 2009, Roznan, Poland, November 6-8, 2009, Re-vised Selected Papers. Springer.
29. Witt, A. & Metzing, D. (2010): Linguistic Modeling of Information and Mark Languages: Contributions to Language Technology. Springer.
30. Zhu, R. (2012): Information Engineering and Applications. Springer.

الفصل الثَّاني

التَّحليل الصَّوتيّ

(الدِّراسة التَّحليليَّة للصَّوت اللُّغَوِيّ في العصر الحديث)

د. أحمد راغِب

١- تحليل الصَّوت اللُّغويّ.

٢- آليات التنفيذ، وإعداد قاعدة البيانات الصوتية.

مدخل

يقوم هذا الفصل على تقديم دراسة تحليليَّة للصوت اللغوي، وذلك من خلال الاتكاء على نتائج التحليل التقني لأصوات اللغة، إلا أن تلك الدراسة لم تكن منبتة الصلة عن جهود علماء العربية الذين قدموا وصفًا تفصيلا لأصوات اللغة العربية بغية الحفاظ على النطق العربي من اللحن أو التبديل؛ فلقد كانت «جهود علماء العربية مثل الخليل وسيبويْه وابْن جنّي في دراسة الأصوات اللّغوية يشار إليها دائما في عصرنا على أنها من الإنجازات المتميزة في الدرس اللغوي، وقامت حولها دراسات ليست قليلة».

وسوف يقومُ الباحثُ في هذه الفصل بمعالجة الدراسة التحليلية للصوت اللغوي من خلال محورين رئيسين:

١- أولهما يتناول تحليل الصوت اللغوي.

٢- والآخر يتناول آليات التنفيذ، وإعداد قاعدة البيانات الصوتية.

١- تحليل الصوت اللغوي

١, ١- الصوت اللغوي

ذهب الراغب الأصفهاني إلى أن «الصوت: الهواء المنضغط عن قرع جسمين، وهما ضربان:صوت مجرد عن تنفس بشيء كالصوت الممتد، وتنفس بصوت ما.

والمتنفس نوعان: غير اختياري كما يكون من الجمادات والحيوانات، ونوع اختياري كما يكون من الإنسان، وهو ضربان:

- ضرب باليد كصوت العود وما يجري مجراه.
- ضرب بالفم في نطق وغير نطق.

فالمنطوق منه: إما مفرد من الكلام، وإما مركب كأحد الأنواع من الكلام. وغير النطق : كصوت الناي».

وذهب ابن سينا إلى أن «الصوت تموج الهواء ودفعه بقوة وسرعة من أي سبب كان».

أما ابن جني فقد ذهب إلى أن الصوت اللغوي «عرض يخرج مع النفس مستطيلاً متصلا حتى يعرض له في الحلق والفم والشفتين مقاطع تثنيه عن امتداده واستطالته، فيسمى المقطع أينما عرض له حرفًا، وتختلف أجراس الحروف بحسب اختلاف مقاطعها».

«وقد ثبت علميًّا أن الصوت اهتزازات محسوسة في موجات الهواء، تنطلق من جهة الصوت، وتذبذب من مصانعه المصدّرة له، فتسبح في الفضاء حتى تتلاشى، يستقر الجزء الأكبر منها في السمع بحسب درجة تذبذبها، فتوحي بدلائلها، فرحًا أو حزنًا، نهيًا أو أمرًا، خبرًا أو إنشاءً، صدى أو موسيقى، أو شيئًا عاديًّا مما يفسره التشابك العصبي في الدماغ، أو يترجمه الحس المتوافر في أجهزة المخ بكل دقائقها». وتتم عملية إنتاج الصوت عبر ما يعرف بالجهاز النطقي «الذي يتمثل في «الرئتين ولسان المزمار والرقيقتين أو الحبلين الصوتيين والأنف والفم بأجزائه».

فالصوت يمثل الصورة الواقعية للغة التي يرى «دي سوسير» أنها «كيان عام يضم النشاط اللغوي الإنساني، في صورة ثقافة منطوقة أو مكتوبة، معاصرة أو متوارثة، وبعبارة أخرى كل ما يمكن أن يدخل في نطاق النشاط اللغوي من رمز صوتي أو كتابي أو إشارة أو اصطلاح فهو من قبيل اللغة، التي إما أن تكون في صورة منظمة ذات قواعد وقوانين، وذات وجود اجتماعي فيطلق عليها اللسان، وهي اللغة المعينة التي تتخذ موضوعًا للدراسة مثل العربية أو الانكليزية وإما أن تكون ذات صورة عامة في العقل الجمعي ويطلق عليها اللغة».

وفي المعجم الوسيط: «الصوت كل ما يسمع، وسببه تموج الهواء نتيجة لقرع أو قلع أو نحوهما» [مادَّة: ص و ت]، ويرى الدُّكتور محمد أحمد كامل، في كتابه (العلوم وحياة الإنسان) أنَّ «الإحساس في الأذن بالطاقة الميكانيكية الناتجة من اهتزاز الأجسام في مجال التردد من ٢٠ إلى ٢٠٠٠٠ ذبذبة في الثانية، ويصدر الصوت عند اهتزاز الأجسام نتيجة لتأثرها بالطاقة الحركية أو المغناطيسية أو الكهربية مثل الشوكة الرنانة أو الآلات الموسيقية، أو اهتزاز الثنايا الصوتية بالحنجرة أو الجرس الكهربي، وتنتقل الموجات الصوتية في الوسط المحيط بالجسم (الهواء مثلا) في صورة كرات من الانضغاط (تتقارب الجزيئات من بعضها) تتبعها كرات من التخلخل (تتباعد الجزيئات عن بعضها)، وتتكون الموجه الواحدة أو الذبذبة من انضغاط كامل + تخلخل كامل، ويعتبر طول

الموجة Wavelength هو طول الذبذبة الواحدة، أو هو المسافة بين مركزين متضاغطين متتاليين أو مركزين متخلخلين متتاليين» .

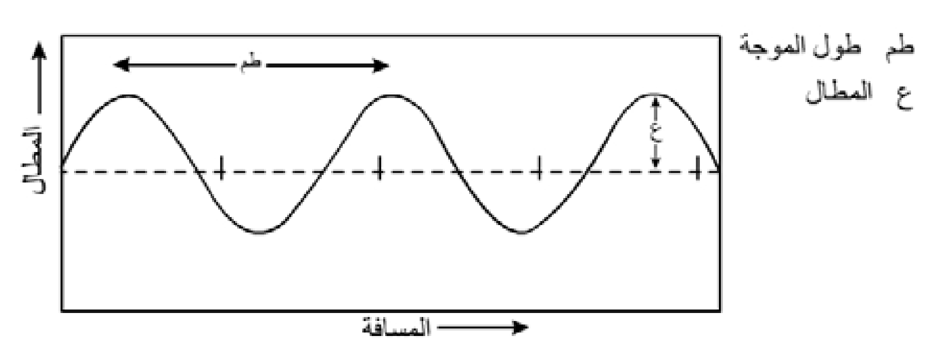
١ , ٢ - سرعة الصوت (Speed of sound)

هي سرعة انتقال الطاقة الصوتية في الوسط، وهي ثابتة في الوسط الواحد بصرف النظر عن نوع الصوت وتردده، ولكنها تختلف من وسط إلى آخر فتتغيَّر طبقا لكثافة الوسط وإعاقته للصوت ودرجة الحرارة، وخصائص الوسط التي تحدد سرعة الصوت هي الكثافة وقابلية الانضغاط، والكثافة هي مقدار الكتلة الموجودة في وحدة الحجم من المادة، وتقيس قابلية الانضغاط مدى سهولة كبس المادة في حيز ضيق. وكلما زادت الكثافة وزادت قابلية الانضغاط، قلت سرعة الصوت. وتزداد سرعة الصوت في السوائل عن الغازات وفي الأجسام الصلبة عن السوائل؛ وذلك لتقارب الجزيئات بها، فالسوائل والأجسام الصلبة بصفة عامة أكثر كثافة من الهواء، وهي أيضًا أقل من الهواء بكثير في قابلية الانضغاط. ولذلك، فإن الصوت ينتقل بسرعة أكبر خلال السوائل والأجسام الصلبة.

١ , ٣- الموجة الصوتية (Sound wave)

هي أحد أشكال انتقال الطاقة، وتتحرك الموجات في وسط مادي[١]، حيث تنتقل فيه الموجات وتنقل الطاقة من مكان إلى آخر بدون إزاحة جزيئات الوسط بشكل دائم، أي أنه لا تنتقل أي كتلة مع انتقال الموجة، ولكن جزيئات الوسط تتحرك بشكل متعامد أو مواز لاتجاه حركة الموجة حول موقع ثابت. وللموجات صفة الدورية، فتكون عادة تكرارا لنمط ما من الشدة في فترات زمنية متتابعة، ويسمى عدد الموجات المارة في مقطع ما مقسوما على وحدة الزمن التردد. وتسمى المسافة الأفقية التي تقطعها الموجة الواحدة طول الموجة، وهو أصغر جزء متكرر مكون للموجات، ويساوي المسافة بين قمتين أو قعرين متتابعين، أو هو المسافة المحصورة بين ٣ قيم صفرية متتابعة للموجة، ويوضح الشكل (٢-١) رسما تخطيطيا لطول الموجة.

١- باستثناء الموجات الكهرومغناطيسية وبعض أشكال الجزيئات الكمّية ذات الخصائص الموجية.

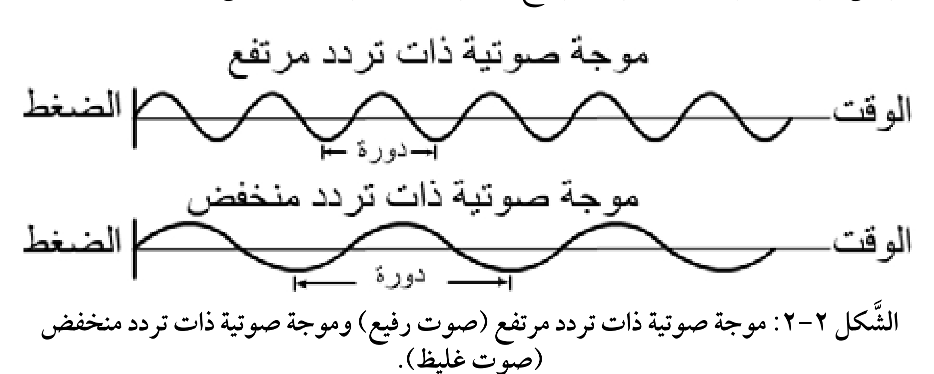


الشَّكل ٢-١: رسم تخطيطي لطول الموجة.

وصفة الدورية للموجة عبارة عن التردد أو التواتر، وهو مقدار تكرر الموجة الواحدة ذات الطول الموجي المتفق عليه في كل وحدة زمن.

وتكون درجة الصوت أعلى كلما كانت الذبذبات أسرع، وكلما زادت سرعة هذه الذبذبات كان الصوت دقيقًا أو حادًّا، ومن ثم فإن الصوت يكون سميكًا إذا قل عدد الذبذبات في الثانية الواحدة. «فالفرق بين شوكة رنانة ذات درجة صوتية عالية وأخرى ذات درجة صوتية منخفضة أن الأولى تعمل عددًا أكبر من الذبذبات في الثانية الواحدة». ويقاس التردد بعدد الذبذبات التي تحدث في الثانية الواحدة».

وإذا كان تردد الصوت (Sound Frequency)[1]: هو عدد الذبذبات أو الموجات الكاملة في الثانية الواحدة فإنه يمكننا رصد هذا التردد كما في الشكل (٢-٢) الذي يعرض لموجة صوتية ذات تردد مرتفع وأخرى ذات تردد منخفض.



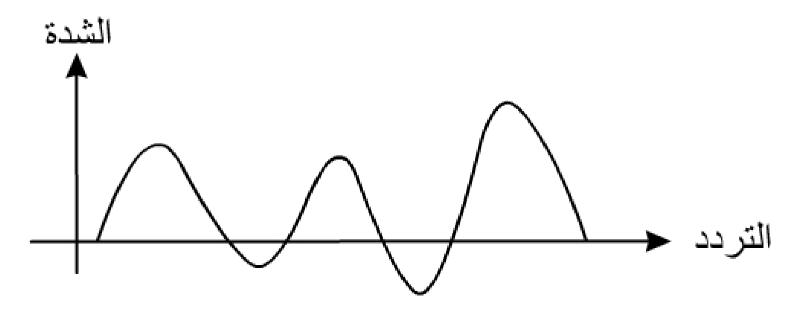
الشَّكل ٢-٢: موجة صوتية ذات تردد مرتفع (صوت رفيع) وموجة صوتية ذات تردد منخفض (صوت غليظ).

١- وحدة تردد الصوت هي هيرتز (Hertz أو Hz أو ذبذبة/ ثانية).

ونلاحظ هنا أن التردد يتناسب عكسيا مع طول الموجة، أي كلما زاد التردد قصر طول الموجة، أي: التردد ٤٠٠٠ هرتز له طول موجة أقصر من التردد ٥٠٠ هرتز.

ويعتبر التردد من العناصر التي تؤدي إلى إدراك بعض الأصوات وتمييزها؛ حيث «يؤدي اختلاف تردد موجات بعض الأصوات اللغوية إلى تغير في إدراكها، فالصوت / س/ ذو تردد عال يفوق ٤٠٠٠ هرتز، فإذا انخفض تردده ليقترب من ٢٥٠٠ هرتز فإن السامع يدركه / ش/ ».

ويمكن رصد العلاقة بين التردد والشدة من خلال الشكل (٢-٣) الذي يعرض رسما تخطيطيا لرصد العلاقة بين التردد والشدة لموجة مفردة.



الشَّكل ٢-٣: رسم تخطيطي يرصد العلاقة بين التردد والشدة لموجة مفردة.

١ , ٤ - التّردد الأساسي

هو أحد العوامل التي «يمكن عن طريقها إنتاج أصوات يختلف الواحد منها عن الآخر[1]. وحين تسمع أي صوتين يمكن أن تقارنهما من هذه الجوانب المختلفة، مثل شوكة رنانة وأرغن، فهما يصدران صوتين مختلفين نتيجة عامل أو أكثر».

ويعبر شكل الموجة عن قيم شكل الموجة أو التردد الأساسي لهذا الصوت، فالصوت ينظر إليه عادةً على أنه عمليّة فيزيائية مادّيّة تتكوّن من جزأين أساسيين:

- منتج حقيقي للصوت؛ ويمثله الحبلان الصوتيان.

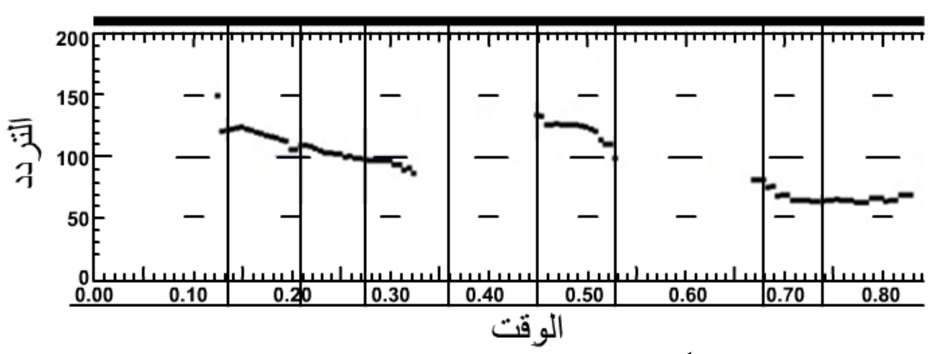
١- بمعنى أنه يمكن عن طريقها تمييز صاحب الصوت مع تشابه الصوتين.

- مشكل للصوت؛ وهو عبارة عن منتج مساعد، يقتصر دوره على إتمام عملية التوجيه واكتساب الصفات التمييزية، وتمثله باقي أعضاء النطق مثل: اللّسان والشّفاه والأسنان... إلخ.

وتحاول عملية تحليل التنغيم رصد هذا التردد الأساسي للصوت المنطوق.

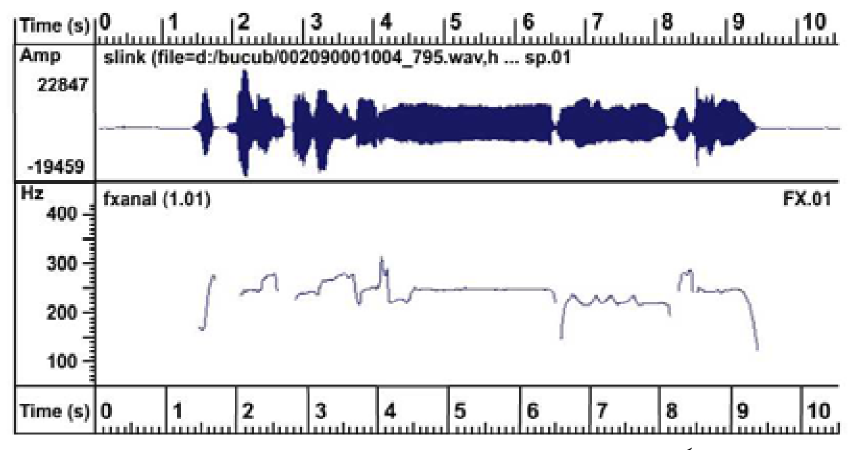
وهناك عدّة قواعد لمعرفة التردد الأساسي تخلص إلى أنه العنصر الأقوى الذي يجعلنا ندرك علاقة التلازم بين الصوت وصاحبه، فبمجرد أن نسمع جملة نعرف قائلها بسبب معرفتنا لتنغيمه. «ويمكن أن نميز صوتًا عن صوت آخر بشكل الموجة التي تُنْتِجُ كلاًّ منهما، وليس شكل الموجة مقياسًا بالمعنى الدقيق، ولكنه مجرد وسيلة لتمييز الأصوات بتحليلها إلى الموجات التي تتكون منها».

وتظهر معالم التردد الأساسي على هيئة منحنيات أفقية موازية لخط الزمن، وتكون أكثر وضوحًا مع الأصوات المجهورة؛ حيث يهتزُّ الحبلان الصّوتيّان، وتتواجد القيم لهذه الحزم بين ٨٠-٢٠٠ هرتز بالنسبة للمتحدثين الذكور، وتتزايد هذه القيمة مع الإناث لضيق الحنجرة عن حنجرة الذكور، وقلة سمك الأحبال الصوتية؛ فتبلغ قيمة هذه الحزم مع الأنثى البالغة ١٥٠-٣٥٠ هرتز. ويوضح الشكل (٢-٤) منحنى التردد الأساسي لكلمة مفردة.



الشَّكل ٢-٤: منحنى التردد الأساسي لكلمة مفردة.

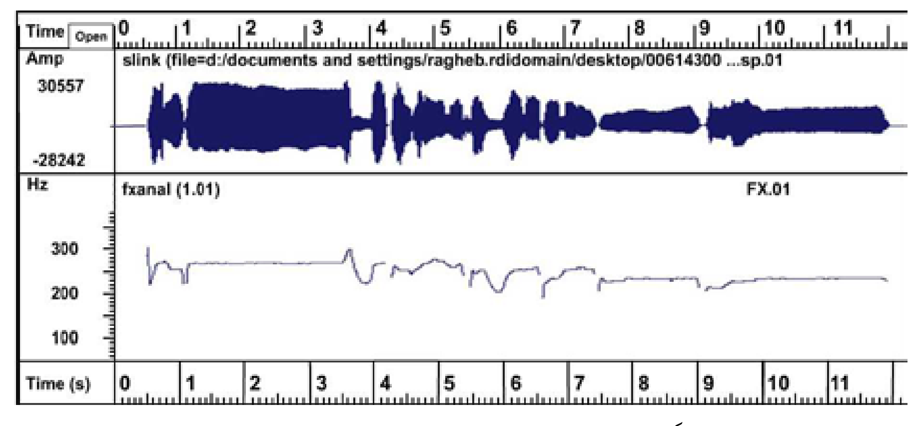
في هذه الصورة تتواجد قيم التردد الأساسي بين ١٠٠ و ١٥٠ هرتز، والصورة عبارة عن رسم طيفي لكلمة «فوناتيك» بصوت ذكر بالغ، ونلاحظ أن التردد الأساسي في المقطع الأول أعلى منه في المقطعين الثاني والثالث، ويظهر الشكل (٢-٥) منحنى التردد الأساسي لجملة خبرية.



الشَّكل ٢-٥: منحنى التردد الأساسي لجملة خبرية.

ويعرض الشكل صورة طيفية لجملة خبرية تمثل النغمة الهابطة؛ حيث يبدأ فيها التردد الأساسي منخفضًا ثم يزداد في منتصف الجملة ليصل إلى أقل انخفاض في نهاية الجملة، والجملة هي قوله تعالى:﴿بِئْسَمَا اشْتَرَوْا بِهِ أَنْفُسَهُم﴾ [البقرة/ ٩٠].

بينما يعرض الشكل (٢-٦) منحنى التردد الأساسي لجملة استفهامية.



الشَّكل ٢-٦: منحنى التردد الأساسي لجملة استفهامية.

ويعرض الشكل (٢-٦) صورة طيفية لجملة استفهامية تمثل النغمة الصاعدة، حيث نلمح فيها أنها لم تحتوِ على ذلك الانخفاض الذي رأيناه في نهاية منحنى التردد الأساسي في الشكل السابق، والجملة هي قوله تعالى:﴿قُل آلذَّكَرَيْنِ حَرَّمَ أَمِ الأنثَيَيْنِ﴾[الأنعام/ ١٤٣].

١,٥ - شدة الصوت (Sound intensity)

هي قياس كمية الطاقة الصوتية لصوت ما بالنسبة إلى كمية طاقة صوتية ثابتة تستخدم كمرجع، ويتم القياس بطريقة لوغارتمية، وتسمى وحدة شدة الصوت «الديسيبل» (decibel) واختصارها dB. ومن أهم أنواع الديسيبل:

ديسيبل ضغط الصوت (dB sound pressure leveldBspl).

- المرجع الثابت هو طاقة صوتية قدرها ٠,٠٠٠٢ داين/ سم٢.
- وتحسب شدة الصوت كالآتي = ٢٠ لو (الطاقة الصوتية المراد قياس شدتها ÷ الطاقة الصوتية المرْجِع).

فمثلا إذا كانت طاقة صوتية قدرها٠,٠٢ داين/ سم٢ فإن شدة الصوت بها تكـــون:

٢٠ × لو (٠,٠٢ ÷ ٠,٠٠٠٢) = ٢٠ لو ١٠٠ = ٢٠ × ٢ = ٤٠ ديسيبل.

حيث: الداين: هو وحدة قياس الضغط [١١]. ومقداره ١/ ١٠٠٠ من الجرام، «وضغطة موجة ما هو عبارة عن عدد الداينات التي تضغط على مساحة سنتيمتر مربع واحد» [١١].

والديسيبل هو وحدة شدة الصوت وهو وحدة ليس لها تمييز؛ لأنها نسبة بين كميتين من الطاقة، وهو الأساس لقياس كل الأصوات في الطبيعة -كلام، ضوضاء، أصوات حيوانات وغيرها- ويتم قياسه عن طريق جهاز قياس شدة الصوت.

ويمثل الديسبل «مقياسًا عمليًّا لقوة الصوت وضغطه، والديسيبل ١/ ١٠ بيل، والبيل هو ضغط أضعف صوت تدركه الأذن الإنسانية» [١١].

وتقاس شدة الصوت «بأجزاء من عشرة آلاف مليون مليون من الوات في السنتيمتر المربع، وتبلغ قوة أضعف صوت تسمعه الأذن جزءا واحدا من عشرة آلاف مليون مليون وات سنتيمتر مربع» [١١].

١,٦ - الرَّسم التَّذبذبيُّ (شكل الموجة)

يعرف الصوت فيزيائيًّا بأنه سلسلة لتغييرات الضّغط في الوسط بين المرسل والمستقبل. فالصوت هو: «اضْطِّرَابٌ مَاديٌّ في الهواء يَتَمَثَّلُ فِي قُوَّةٍ أَوْ ضَعْفٍ سَرِيعَيْنِ للضَّغْطِ المَتحَرِّكِ مِنَ المصْدَرِ في اتجاه الخارج، ثم في ضَعْفٍ تَدْرِيجِيٍّ إِلِى نُقْطَةِ الزَّوَالِ النِّهَائِيّ» [٢٥، ٢٦]. فالمرسل عندما يبدأ عملية إصدار الصوت فإن الصوت الخارج منه يقوم بالضغط على الهواء في شكل موجات متلاحقة تصطدم بالمستقبل ميكرفون أو أذن مثلًا- فتحدث عملية التخلخل التي تصاحب الضغط، وأكثر الأمور شيوعًا لتمثيل معالم الصوت الحسية ورصدها هو الرّسم التّذبذبيّ، الذي يطلق عليه الشكل الموجي. ويمثل المحور الأفقي فيه محور الزّمن وتتم قراءته من اليسار إلى اليمين بينما يمثل المنحنى الرأسي قوة هذا الصوت في لحظة معينة، في الوقت الذي تظهر فيه عدد الترددات الذبذبية للحظة الواحدة على هيئة تتابع وتلاحق هذه الخطوط الرأسية لهذا المنحنى.

فعندما ننطق أية كلمة (دفقة صوتية واحدة) في البداية سيتم التنفس من الرئتين ثم يخرج الهواء إلى الحنجرة الذي قد يسبب عند مروره بها اهتزاز الحبلين الصوتيين ويستمر في مسيره حتى يصل إلى القناة الصوتية وينتقل منها إلى التجويف الفموي[١]، وبعدها تخرج الكلمة وتصبح مسموعة وتتحول إلى شكل الموجة. ويظهر الشكل (٢-٧) صورة للرسم التذبذبي من خلال برنامج Cool Edit Pro.



الشَّكل ٢-٧: صورة للرسم التذبذبي (شكل الموجة)، كما يعرضه برنامج Cool Edit Pro.

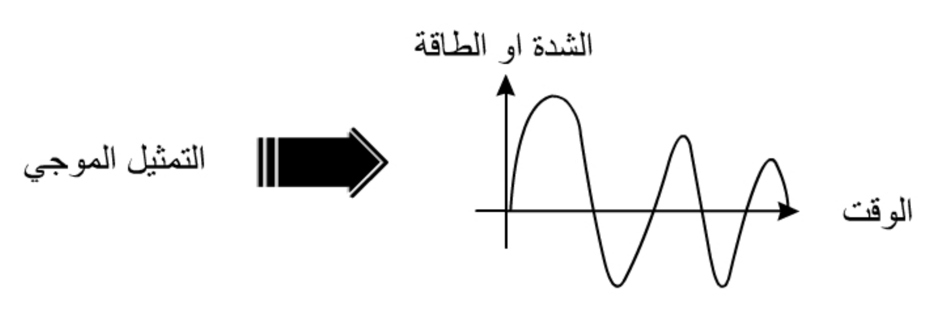
من الشكل (٢-٧) يمكن تحليل صوت الفاء بأنه صوت استغرق ١٢٦ ميللي ثانية[٢]، وهو صوت عالي التردد ٤٠٠٠ هرتز، ضعيف الإسماع. بينما تمثل النون قوة

١- وفي الأصوات الأنفية يضطر الهواء للمرور بالتجويف الأنفي (nasal cavity).

٢- كلمة ميل Milli تعني ١/ ١٠٠٠، ويعني ميللي ثانية ١/ ١٠٠٠ من الثانية وفي الأصوات الأنفية يضطر الهواء للمرور

الإسماع ٢٥٠٠ هرتز، مع ترددات متوسطة ٦٨٠ ذبذبة، وقد استغرق هذا الصوت ٥٥ ميللي ثانية.

أما صوت الدال فهو منخفض الترددات، ومنخفض الإسماع، وقد استغرق ١٣٠ ميللي ثانية. وبالنظر إلى الشكل (٢-٨) نرى إحداثيات شكل الموجة من خلال المحورين السيني والصادي.



الشَّكل ٢-٨: وتظهر فيه إحداثيات شكل الموجة حيث المحور السيني هو الوقت والصادي يمثل الشدة.

١,٧ - علم الأصوات وطرق التحليل الصوتي

الفوناتيك هو جزء من علم اللغة، إنه معني بدراسة الحدث الناتج بواسطة الأعضاء الصّوتيّة البشريّة، ويعنى بالأخص بدراسة الحدث اللغوي الذي يُسْتَخْدَم في الكلام البشري، ويعد التحليل الطيفي للأصوات من أهم جوانب البحث الصوتي، وهو الجانب الذي يُشَار إليه كثيرًا بالجانب الصوتي التجريبي، أو التحليل الميكني للصوتيّات البشرية.

ويتم التّحليل الصوتي (Phonological analysis) بواسطة أداة أو أكثر من الأدوات المتاحة لهذا الغرض، ومن بينها أفلام أشعة إكس (X_Ray)، وأنابيب جريان الهواء (Pneumotachogrph)، والكيموجراف (Kymograph)، والإليكتروميوجراف (Electromyograph)، والإسبيكتروجراف (Spectrograph)، واللارينجوجراف (Laryngograph)، ورسام الحنك الإلكتروني (Electro palato graph).. إلخ.

بالتجويف الأنفي (nasal cavity).

وهدف كل هذه الأجهزة والأدوات إظهار المعالم الحسية للصوت بشكل ما، بالإضافة إلى محاولة رصد بعض جوانب تلك الإشارات أو المعالم الحسية للصوت البشري على هيئة صورة يمكن طبعها على الورق أو رصدها على شاشة كمبيوتر.

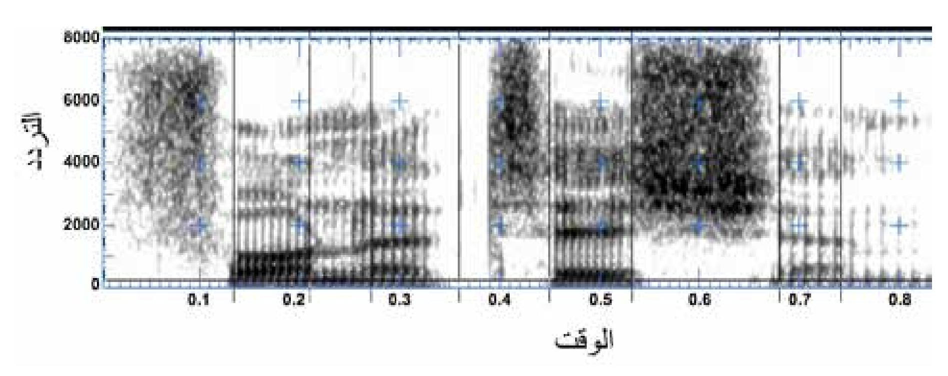
والكمبيوتر اليوم هو من أكثر الأدوات سهولة واستخدامًا؛ فعن طريقه تعد عمليَّة التَّحليل الصوتي أبسط كثيرًا وأسرع عادةً إذا ما قورنت بالأدوات الأخرى، مع الحفاظ على جودة النتائج، وإمكانية جمع كل الأدوات في برنامج واحد.

١ ,٨ – الصورة الطيفية (Spectrogram)

نستطيع تمثيل الصوت اللغوي عن طريق الصورة الطيفية وهي ذات ثلاثة أبعاد: الوقت، والتردد، والشدة، ويمثل المحور الأفقيُّ في الصُّورة الطَّيفيَّة منحنى الزَّمن بالميللي ثانية، ويمثل التردد المحور الرأسي بالهيرتز، أما البعد الثّالث وهو مدى التّردّد أو شدة الصوت فيُمَثَّله قرب أو بعد لون الصورة عن لون الخلفية. وقد تعتبر الصّورة الطّيفيّة عددا من السّلاسل المتوالية إذا نظرنا إليها من أعلى (مسقط أفقي)؛ حيث تُمَثَّل الزّيادات أو القمم الرأسية في الصور الطيفية بالبقع المظلمة في الصّورة الطّيفيّة، وكلتاهما تعبر عن شدة الصوت المنطوق.

ويعد جهاز المطياف من أفضل الأجهزة التي خدمت الصوتيات الأكوستية إن لم يكن أفضلها على الإطلاق، حيث يقوم بعرض صورة مرئية للنطق تكشف عن خصائص الكلمة المنطوقة، وهو يفيد في معرفة صفات الأصوات اللغوية وترجيح أوجه الخلاف في وصف بعض الأصوات التي دار حولها الجدل قديمًا وحديثًا، «ونظرًا لأن موجات الصوت اللغوية من النوع المركب فإن عرضها باستخدام جهاز عارض الذبذبات لا يقدم كل التفاصيل عن الموجة الصوتية. بينما يقدم المطياف ثلاثة أبعاد للموجة المرسومة وهي: التردد، والشدة، والزمن. وهذا يعين الباحث في معرفة زمن الصوت، والتردد الأساسي، والنطاق الرنيني وشدته؛ ولذا فإن معظم دراسات أكوستية الصوت اللغوية تعتمد على المطياف بشكل أو بآخر» [٢١].

وبالنظر إلى الشكل (٢-٩) نرى صورة طيفية مأخوذة من المطياف توضح الأبعاد الثلاثية للصوت المنطوق.



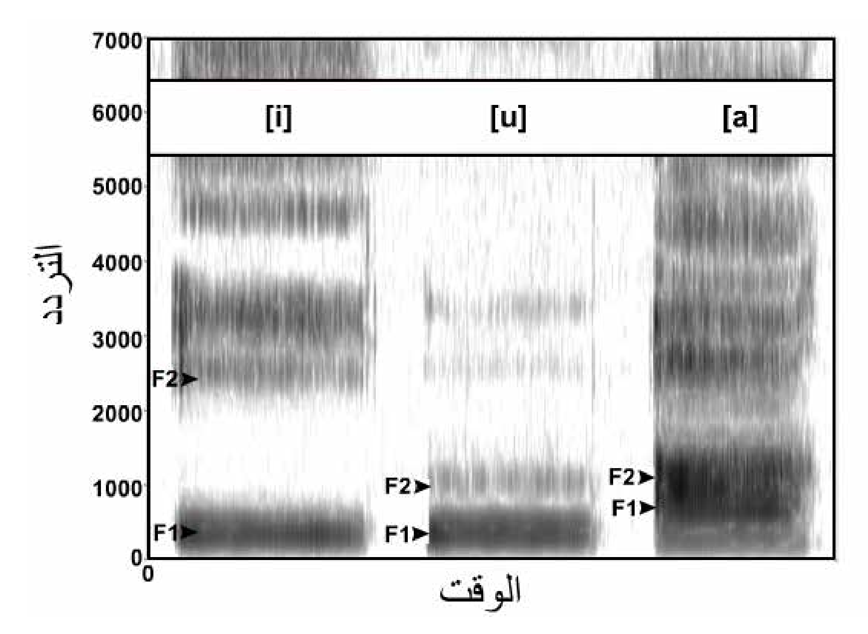
الشَّكل ٢-٩: صورة طيفية مأخوذة من المطياف توضح الأبعاد الثلاثية للصوت المنطوق.

من الصّورة يتضح لنا اختلاف طبيعة الأصوات المنطوقة حسب معطيات الصورة الطيفية، ففي الحرف الاحتكاكيّ المهموس نجد أن الطّاقة مركّزة بكثرة في نطاق التّردّد أو الحزم، وفوضويّ إلى حدّ ما (شبيه بالضّوضاء) في مظهره. أما في باقي الأصوات المهموسة-الانفجارية على سبيل المثال-فإن هذه الأصوات تتكوّن من فترة صامتة حتّى تظهر الطّاقة فجأة وبقوة في نطاقات التّردّد أو الحزم، على هيئة انفجار.

وتبدو الأصوات المجهورة أكثر تنظيمًا، وتبدو أعلى السّلسلة (البقع المظلمة) في الواقع على هيئة خطوط في وضع أفقيّ الشّكل عبر الصّورة الطّيفيّة، وتمثّل هذه الحزم الذّبذبات؛ حيث يعطي شكل الفم رنين الصوت، هذه الحزم تُسَمَّى المعالم، ويتم ترقيمها من الأسفل إلى الأعلى على هذا النحو ف١ ف٢ ف٣ أو F1 ،F2 ،F3 إلخ؛ وتختلف مواضع هذه الحزم باختلاف نوع الصوت المنطوق، ومن الممكن تحديدها بشكل تقريبي مع كل فونيم.

فهذه الحزم الصوتية التي يطلق عليها المعالم هي «الترددات أو مجموعة الترددات التي تشكل نوع الصوت وتميزه عن الأصوات الأخرى ذات الأنواع المختلفة» [٢٠].

ومن خلال الشكل (٢-١٠) نرى صورة طيفية توضح أبعاد الحزم الصوتية للحركات القصيرة.



الشَّكل ٢-١٠: صورة طيفية توضح أبعاد الحزم الصوتية للحركات القصيرة، مصدر الصورة: موسوعة ويكيبيديا.

١, ٩- الكتابة الصّوتيّة

تمثل الكتابة العادية طريقة واحدة للتعبير عن الأصوات عن طريق مجموعة من الرموز الكتابية. فإذا كانت اللغة عبارة عن «أصوات يعبر بها كل قوم عن أغراضهم» فإن الكتابة هي رموز يعبر بها كل قوم عن أصواتهم، وتقوم الكتابة على نظام رمزي صوتي، «ومهمة هذا النظام أن يزودنا بالرموز التي يمكن أن نعبر بها عن القيم الصوتية المختلفة للمواد اللغوية في لغة ما» [٢٢].

«فمن الخصائص العالمية للغة التناظر بين الحرف والصوت، بين الهجاء والنطق، حيث تماثل تلك العلاقة نظام الكتابة الصوتية العالمي وهو النظام النموذجي العالمي الذي توافق عليه علماء اللغة والأصوات» [٣٢]، وقد عدَّ سابير اللغة العربية من اللغات الخمس الوحيدة الحاملة للحضارة الإنسانية والتي يسهل إدراجها في نظام الكتابة الصوتية العالمية.

ومنذ فترة ليست بالبعيدة استطاع علماء الأصوات أن يبتكروا ألفبائية جديدة للتعبير عن الأصوات المختلفة. وتحاول هذه الألفبائية الحفاظ على قرب العلاقة بين الرموز الكتابية والصّوت المنطوق. وتعتبر الألفبائية المستخدمة على نطاق واسع هي: IPA أو (International Phonetic Association). وهو اسمُ جمعيّة صوتيّة دوليّة نشأت أواخر القرن التّاسع عشر، وكانت المراجعة الأخيرة لها في عام ١٩٨٩م. ويعمل نظام IPA على إيجاد رمز لكلّ فونيم في كلّ لغة بشريّة، ووضع علامات تشكيلية للتّغيّرات التي تعتري أية لغة من اللغات.

وقد قسمت هذه الجمعية اللغوية الأصوات الإنسانية إلى ستة أقسام، هي:

- رموز لأصوات صامتة رئوية.
- رموز لأصوات صامتة غير رئوية.
- رموز لأصوات صائتة (الحركات).
- رموز لأصوات نغمية.
- رموز لعلامات تمييزية.
- رموز الفوقطعيات.

ثم تطورت فكرة هذه الألفبائية، واكتست ثوبًا جديدًا يتوافق مع سهولة التعامل مع الحاسب الآلي؛ بحيث يجد كل رمز صوتي مكانًا له على لوحة المفاتيح بالجهاز، وقد تم ذلك بفضل جهود نظام سامبا اللغوية الصوتية

(SAMPA – Speech Assessment Methods Phonetic Alphabet)

فجمعت الرموز الصوتية العربية على النحو المبين في الجدول (٢-١):

الرمز	الكلمة المثال	الكتابة الصوتية
الصوامت الاحتكاكية:		
F	فيل	fi:l
V	نفين	nivi:n
T	ثلاث	Tala:T
D	ذكر	Dakar
D`	ظلام	D`ala:m
S	سعيد	sa?`i:d
Z	زميل	zami:l
s`	صغير	s`aGi:r
S	شمس	Sams
Z	جميل	Zami:l
X	خطاب	xit`a:b
G	غرب	Garb
X\	حلم	X\ilm
?` (?\)	علم	?`alam
H	هواء	hawa:?
الصوامت الأنفية:		
M	مال	ma:l
N	نور	nu:r
الصوت التكراري:		
R	رمال	rima:l
الصوتان الجانبيان:		
L	لا	la:
l`	الله	?al`l`ah

الرمز	الكلمة المثال	الكتابة الصوتية
الحركات:		
I	ظل	D`il
A	حل	X\al
U	عمر	umr`?
i:	عيد	i:d`?
a:	مال	ma:l
u:	فول	fu:l
أنصاف الحركات		
W	واحد	wa:hid
J	يوم	Jawm
الصوامت:		
الصوامت الانفجارية:		
B	باب	ba:b
T	تسع	tis?`
D	دار	da:r
t`	طابع	t`a:bi?`
d`	ضرب	d`arab
K	كبير	kabi:r
G	جميل	gami:l
		في العامية المصرية
?	أكل	akl?
Q	قلب	Qalb
P	برس	Paris

الجدول ٢-١: الأصوات العربية.

٢- آليات التنفيذ، وإعداد قاعدة البيانات الصوتية

يمكن تفصيل نقاط البحث في هذا المبحث إلى:

٢, ١ - أجهزة التحليل الصوتي

٢, ٢ - الأدوات والبرامج التحليلية الحاسوبية

٢, ١ - أجهزة التحليل الصوتي

اعتمدت الدراسات الصوتية في التراث العربي على قدر كبير من الملاحظة المباشرة والتجارب الذاتية؛ إلا أنه تماشيًا مع الثورة الصناعية وما نتج عنها من فرض للمنهج التجريبي في شتى مجالات البحث العلمي حدث تطور كبير في آليات معالجة هذه الدراسة «واستخدم علم الأصوات منذ القرن التاسع عشر الأجهزة الدقيقة سواء في التسجيل أو التحليل، وتعاونت أقسام الصوتيات في مختلف الجامعات مع أقسام الفسيولوجيا والفيزياء والهندسة الكهربائية ومعالجة الكلام وطب الأسنان وغيرها» [٢٠]. فقد أصبحت الصوتيات علمًا معمليًّا «يعتمد على التجارب التي بدورها تعتمد على معامل مجهزة تجهيزًا حديثًا يواكب التطور العلمي والتقني الذي نعيشه، وهناك أجهزة تخدم فروع الصوتيات الثلاثة: النطقية والأكوستية والسمعية، فيمكن متابعة حركات الجهاز الصوتي وعضلاته والهواء المنساب داخله، إضافة إلى ذلك فإنه يمكن وضع تجارب واستخدام أجهزة دقيقة؛ لمعرفة الكيفية التي تتم بها عمليتي السمع والإدراك، كل هذا الكم من الأجهزة يجعل النتائج التي يخرج بها الباحثون في علم الصوتيات phonetics دقيقة لدرجة كبيرة وتتمتع بالثبات والموضوعية» [٢١].

وسوف يعرض الباحثُ فيما يلي لأهم الأجهزة المستخدمة في التحليل الصوتي:

- **جهاز الرسم الحنجري (Laryngograph)**

«وهو جهاز إلكتروني يمكننا من استنتاج حالتي الفتح والغلق للأوتار الصوتية عن طريق تسجيل اتجاه التيار من أحد جانبي الحنجرة إلى الجانب الآخر، ويمكن تحويل هذا التسجيل إلى صوت يمثل نتيجة عمل الأوتار الصوتية دون تأثير أي رنين صادر عن القناة العليا، كما لو كنا قد فصلنا تجاويف ما فوق الحنجرة، وسمعنا ذبذبة الأوتار الصوتية بدونها» [٢٠].

- **مقياس التنفس (Spirometer/Respirometer)**

يفيد هذا الجهاز في قياس مقدار وقوة واتجاه هواء الزفير، والجهاز عبارة عن خرطوم مطاطي متصل بجهاز قياس، حيث يوضع الأنبوب المطاطي في فم المتحدث مع إغلاق فتحتي الخيشوم، فيقوم الجهاز برصد مقدار وقوة واتجاه التنفس.

- **مقياس انسياب الهواء (Pneumotachogrrph)**

فائدة هذا الجهاز قياس مدة انسياب هواء الزفير الخارج من الأنف والفم معًا أو كلٍ على حدة، والجهاز عبارة عن كمامة توضع حول الفم والأنف مع وجود عازل بينهما، وتتصل هذه الكمامة بجهاز قياس فترة خروج الهواء.

ويفيد الجهاز في «معرفة حجم الهواء الخارج من الجهاز الصوتي مع كل صوت، وعلاقة حجم الهواء مع الأوضاع المختلفة للجهاز الصوتي. فمن المعلوم أن حجم الهواء الخارج من الجهاز أثناء نطق / ز/ أقل من ذلك المصاحب للصوت / س/ ، حيث إن الرقيقتين الصوتيتين تعيقان انسياب الهواء في الأول فيقل حجم الهواء الخارج، وذلك بعكس ما يحدث أثناء نطق / س/ ، ويمكن بواسطة هذا الجهاز معرفة ما إذا كان الصوت أنفيًّا أم فمويًّا أم أنفمويًّا، وذلك بتتبع حجم الهواء الخارج من الفم والأنف أثناء نطق الصوت موضع الدراسة» [٢١].

- **منظار الحنجرة أو المجهر الحنجري (Laryngoscope)**

يقوم هذا الجهاز برصد حركة الأوتار الصوتية، وهذا المنظار عبارة عن مرآة مستديرة قطرها بوصة، ويوضع هذا المنظار في فم المتحدث فيعكس صورة الأوتار الصوتية، ويبين إذا كان الصوت المنطوق مهموسًا أو مجهورًا. «ويعيب هذه الآلة أنها تتدخل في سير الكلام الطبيعي، وأنها لا يمكن استعمالها إلا في حالات محدودة» [٢٠].

- **مجهار العضلات (Electromyograph)**

يفيد هذا الجهاز في رسم إشارات تعبر عن حركة العضلات المصاحبة لعملية النطق؛ حيث يقوم بقياس الشحنات الكهربائية الصادرة من المخ التي تؤدي إلى تناغم ما يزيد عن مئة عضلة في الجهاز الصوتي؛ لتوليد الصوت الإنساني بشكل سلس.

- **الأشعة السينية (X_ Ray)**

تقوم هذه الأشعة بتصوير أعضاء الجهاز الصوتي أثناء قيامه بعملية إنتاج الكلام، فهذه الأشعة «تسمح بدراسة كل موقع لأي عضو من أعضاء الكلام عند أي نقطة أثناء الكلام، وهناك كذلك الصور المتحركة لأشعة إكس التي تسجل حركات هذه الأعضاء أثناء النطق، ومن الممكن مصاحبة هذه الأفلام بتسجيل صوتي حتى يمكن أن تستمع إلى الصوت وتشاهد الحركات التي تقوم بها أعضاء النطق أثناء إحداث هذه الأصوات» [٢٠]. وقد اتجه العلماء حديثًا إلى استخدام الأشعة السينية المنخفضة لتفادي الأضرار الناتجة عن تعرض الجهاز الصوتي للأشعة السينية لفترة طويلة.

- **رسام الحنك الإلكتروني (Electropalatograph)**

يستخدم هذا الجهاز في رسم الحنك حال نطق الأصوات بطريقة إلكترونية، والحنك هو التركيب الفاصل بين جوف الأنف وجوف الفم، فهو سقف الفم وأرضية الأنف، ويمتد الحنك نحو الخلف ليفصل جزئيًّا بين التجويف الأنفي والتجويف الفموي، للحنك شكل القبة ويتألف من قسمين هما الحنك الصلب الذي يشكل الجزء الأمامي والحنك الرخو الذي يشكل الجزء الخلفي منه، ويمثل الحنك الصلب الجزء العظمي من الحنك ويشكل الثلثين الأماميين منه، ويتألف من الناتئين الحنكيين للعظمين الفكيين العلويين في الأمام ومن الصفيحتين الأفقيتين للعظمين الحنكيين في الخلف، ويفصل الحنك الصلب جوف الفم عن جوف الأنف.

أما الحنك الرخو أو الحنك اللين فهو الجزء الخلفي من الحنك «وهو عبارة عن طية ليفية عضلية متحركة ومعلقة بالحافة الخلفية للحنك الصلب، هذه الحافة الخلفية للحنك الرخو حرة ولها في المستوى الناصف استطالة تدعى اللهاة» .

ويقوم جهاز رسام الحنك الإلكتروني برسم «أجزاء الحنك التي يلامسها اللسان أثناء الكلام كل ١/ ١٠٠ من الثانية = ١٠ ميللي ثانية. يتكون الجهاز من حنك صناعي مصنوع من البلاستيك ويحتوي على ٦٤ حالبًا- يزيد عددها أو ينقص حسب نوع الجهاز، يتم تثبيت الحنك الصناعي في الفم ملاصقًا للحنك الطبيعي وموصلًا بحاسوب. أثناء التجربة، يعرض الجهاز الحوالب التي يلامسها اللسان أثناء التحدث إما على الشاشة أو بالطابعة على الورق» [٢١].

- رسام الحنجرة الإلكتروني (Electrolaryngograph)

«يتكون الجهاز من حالبين يوضعان على سطح الجلد المغطي لغضروف الحنجرة بحيث يقعان على الصفيحتين الغضروفيتين، يتابع الحالبان حركة تقارب الرقيقتين الصوتيتين أثناء إصدار الصوت، وتنقل الإشارة إلى جهاز عرض أو رسم، والجهاز يقيس تردد الرقيقتين الصوتيتين ونمط التردد، وله فوائد في تحديد ما إذا كان عمل الرقيقتين الصوتيتين طبيعي أم لا» [٢١].

٢ , ٢- الأدوات والبرامج التحليلية الحاسوبية

للقيام بعمل قاعدة البيانات يتعين على الباحث الاستعانة ببعض الأجهزة والبرامج الأساسية والأدوات المساعدة للخروج بأفضل عمل ممكن، وهي كالتالي:

COOL EDIT PRO

SFS (Speech Filing System)

EMU SPEECH TOOLS

HTK (Hidden Markov Model Tool Kit)

وفيما يلي إشارة وجيزة عن هذه الآليات وطريقة إفادته منها.

- برنامج (COOL EDIT PRO)

هذا البرنامج من أهم البرامج المستخدمة في إنشاء قواعد البيانات الصوتية؛ وذلك لقدرته على إظهار الصوت على هيئة مرئية، تمكننا من ملاحظة الصوت بدقة، وإجراء التعديلات المناسبة عليه من حذف الزوائد التي تشوب الصوت ولصقها وتقطيعها وإزالتها، وبه يتم تجهيز الصوت تجهيزًا محكمًا حتى يصبح جاهزًا لأغراض البحث الصوتي. ويتعامل هذا البرنامج مع ملفات الصوت الطبيعي وملفات الصوت المضغوط. ويفيد هذا البرنامج في عرض الصورة الطيفية للصوت المسجل، كما يعرض لمستوى ذبذبات كل صوت وشدته ودرجة علوه أو انخفاضه، بالإضافة إلى إمكانية إدراج خلفيات للصوت وقياس صداه. ويعرض الشكل (٢-١١) صورة توضيحية للبرنامج.

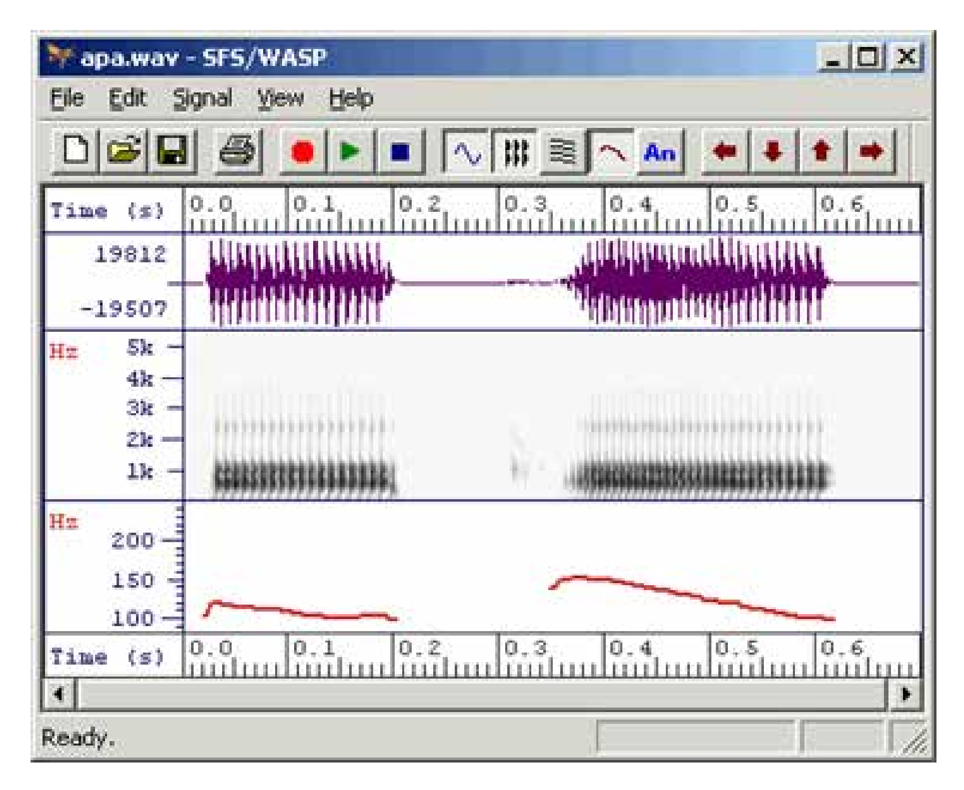


الشَّكل ٢-١١: صورة توضيحية لبرنامج COOL EDIT PRO.

http://www.softpedia.com/get/Multimedia/Audio/Au-dio-Editors-Recorders/Cool-Edit-Pro.shtml

- **برنامج (SFS -Speech Filing System)**

برنامج مستخدم حديثا في مجال تجهيز قواعد البيانات؛ حيث تم الاعتماد عليه مؤخرًا، وذلك في خطة تطوير الأدوات المستخدمة في بناء قواعد البيانات الصوتية، وبالنسبة لهذا البرنامج فقد أظهر كفاءة ملحوظة في مجال تحديد بدايات ونهايات المقاطع الصوتية تحديدًا دقيقًا، كما يتيح لنا هذا البرنامج مجالا خصبا للتعرف على صور مختلفة من مستويات الصوت من حيث الطاقة والقــوة، ومــن حيــث الأداء الصــوتي والحنجري والنبري للكلمات المختلفة. ويعرض الشكل (٢-١٢) صورة توضيحية لبرنامج SFS.

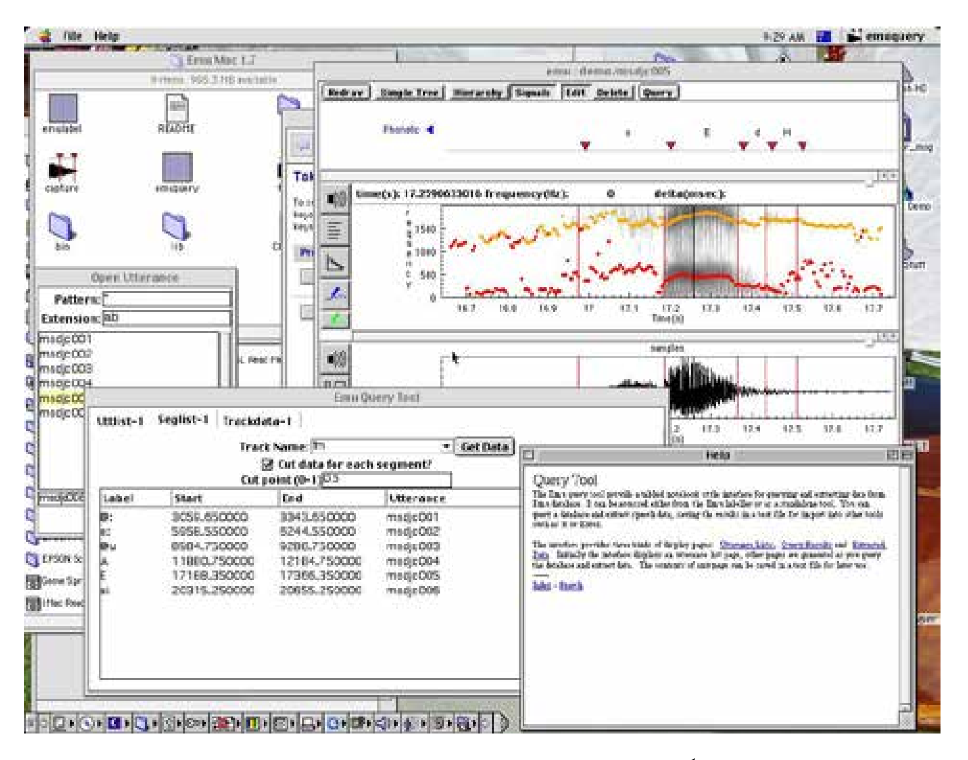


الشَّكل ٢-١٢: صورة توضيحية لبرنامج SFS.
www.phon.ucl.ac.uk/resource/sfs

- **برنامج (EMU SPEECH TOOLS)**

برنامج ترجع أهميته الشديدة هنا إلى أنه يتيح لنا سماع الفونيمات كل على حدة مع ظهور الكتابة الصوتية لكل ملف من الملفات، ويظهر (spectrogram) وكذلك أيضًا (samples)، وهما مَساران مهمان في تحديد مواضع الفونيمات وتحريكها إلى أنسب المواضع، حيث تبدأ الفونيمات من أماكنها الصحيحة، كما أن أي إجراء يتم على هذا البرنامج يكون تعديلا في ملفات lab المهمة، كما يمكننا البرنامج من تغيير أو إزالة أية فونيمات أو إضافتها حتى تكون الكتابة الصوتية مطابقة تماما للصوت.

ويوضح الشكل (٢-١٣) صورة توضيحية لبرنامج EMU.



الشَّكل ٢-١٣: صورة توضيحية لبرنامج **EMU**.

www.emu.sourceforge.net

- **برنامج (Hidden Markov Model Tool Kit- HTK)**

هذا البرنامج عبارة عن عدة خطوات معقدة نسبيًّا تتكون من عدد من البرامج لكل منها أهميتها كجزء رئيسي أو مكمل لخطوات البرنامج، وترجع أهمية هذا البرنامج إلى أنه البرنامج الذي يستطيع أن يقطع الكلمات إلى فونيمات كل على حدة ويحدد بداياتها، كما أن ملفات lab تنتج عن هذه الخطوة، التي نحتاج إليها في إجراء التعديلات على مواضع بدايات الفونيمات، كما تعد تلك الملفات جزءًا رئيسيًّا في قاعدة البيانات التي يتم تخليق الصوت منها [٣٥].

ببليوجرافيا مرجعيَّة

١. ابن جني، أبو الفتح عثمان (١٩٨٥): سر صناعة الإعراب، تحقيق حسن هنداوي، دار القلم، دمشق، ط ١.

٢. ابن سينا، علي الحسين: الرَّئيس: تحقيق محمد حسان الطيان ويحيى مير علم، مطبوعات مجمع اللُّغة العربيَّة بدمشق.

٣. أحمد، أحمد أحمد راغب أحمد (٢٠٠٤): فونولوجيا القرآن - «دراسة لأحكام التجويد في ضوء علم الأصوات الحديث، (رسالة ماجستير)، كلية الآداب، جامعة عين شمس.

٤. أحمد، أحمد أحمد راغب أحمد(٢٠١٣): دور المؤثرات الصوتية في تقدير المدى الزمني للفونيم، مجلة الدراسات اللغوية والأدبية، العدد ٣، المجلد .

٥. أحمد، أحمد أحمد راغب أحمد(٢٠١٥): تأصيل النظرية الصوتية عند إخوان الصفا، مجلة الدراسات اللغوية والأدبية، العدد ٦، المجلد الثاني.

٦. أحمد، أحمد أحمد راغب أحمد (٢٠١٥): قضايا صوتية خلافية في ضوء التحليل الصوتي الحاسوبي، مجلة جامعة القدس المفتوحة، العدد ٣٦، المجلد الأول.

٧. أحمد، أحمد أحمد راغب أحمد (٢٠١٣): الدرس الصوتي العربي بين النظرية والتطبيق، دار تمان إيمان، ط٢.

٨. أحمد، أحمد أحمد راغب أحمد (٢٠١٣): النون والميم في القرآن الكريم دراسة صوتية حاسوبية، دار تمان إيمان، ط٢.

٩. الأصفهاني، الرَّاغب: مُفرَدات القُرآن، تحقيق مُحمَّد سيِّد كيلاني، دار المعرفة.

١٠. أنيس، إبراهيم (١٩٧٥): الأصوات اللغوية، مكتبة الأنجلو، ط ٥.

١١. أيوب، عبد الرحمن (١٩٨٤): الكلام إنتاجه وتحليله، مطبوعات جامعة الكويت.

١٢. بولجرام، أرنست (١٩٧٧م): في علم الأصوات الفيزيقي، ترجمة د. سعد مصلوح، ط.

١٣. الحمد، غانم قدوري (١٩٩٦): أبحاث في علم التجويد، دار عمار للنشر والتوزيع، الأردن.

١٤. الحمد (١٩٨٦)، غانم قدوري: الدراسات الصوتية عند علماء التجويد، مطبعة الخلود ببغداد.

١٥. الخولي، محمد علي (١٩٨٧): الأصوات اللغوية، مكتبة الخريجي، ط ١.

١٦. شاهين، عبد الصبور (١٩٧٧): المنهج الصوتي للبنية العربية، القاهرة.

١٧. الصّغير، محمّد حسين علي (١٩٩٤): الصوت اللغوي في القرآن الكريم، دار المؤرخ العربي، لبنان.

١٨. ضوة، إبراهيم (٢٠٠٠): محاضرات في اللغة العربية والحاسب، دار الثقافة العربية، ط١.

١٩. العاني، سلمان حسن (١٩٨٣): فونولوجيا العربية، ترجمة ياسر الملاح، مطبوعات النادي الأدبي الثقافي بجدة، ط١.

٢٠. عمر، أحمد مختار (٢٠٠٠): دراسة الصوت اللغوي، عالم الكتب.

٢١. الغامدي، منصور بن محمد (٢٠٠٠): الصوتيات العربية، مكتبة التوبة، ط ١.

٢٢. فتيح، محمد (١٩٨٤): الأصوات العامة والأصوات العربية، دار الثقافة العربية، القاهرة.

٢٣. قدور، أحمد محمد (١٩٩٨): أصالة علم الأصوات عند الخليل من خلال مقدمة كتاب العين، دار الفكر المعاصر، بيروت، ط١.

٢٤. مصلوح، سعد (١٩٨٠): دراسة السمع والكلام، القاهرة.

٢٥. الموسوي، مناف (١٩٩٨): علم الأصوات اللغوية، عالم الكتب، بيروت، ط١.

٢٦. هلال، عبد الغفار حامد (١٩٨٨): أصوات اللغة العربية، مكتبة الأنجلو المصرية، ط٢.

27. Ahmed Ragheb Ahmed Mahmoud (2015). The acoustic efforts of the Medieval Arab philosophers: brethren of purity as a model. World Journal of Social Science Research (WJSSR), 2 (2). , 0 pp. 191-199. ISSN 2332-5534.

28. Ahmed Ragheb Ahmed Mahmoud, Muhammad Sabri Sahrir, Rahmah Ahmad H. Osman (2013). Integration of an interactive program in learning Arabic language for non-native speakers via virtual tutor. GEMA Online Journal of Language Studies, 13 (3). , 0 pp. 117-131. ISSN 1675-8021.

29. Gleson, H. A. (1961). (TR). An Introduction to Descriptive Linguistics Reued, New York.

30. Hartmann, R. & Stark, F. (1973). Dictionary of Language and Linguistics, London.

31. Ikoyo-Eweto, E ; Ekiugbo, Philip, A Phonetic analysis of Uvwiẹ vowels, The Journal of West African Languages, 2017, Vol.44(2), p.1.

32. Loo, Alfred ; Chung, C. W. ; Lam, Alan, Speech Analysis and Visual Image: Language Learning, Gifted Education International, 2016, Vol.32(2), p.100-112.

33. Robins, R. H. (1980). General Linguistics an Introductory Survey, Third Edition.

34. Sapier, E. D. (1921). Language: an Introduction to the study of Speech, New York, Harcourt, Brace and company.

35. Saussur, F. D. (1959). Course in general Linguistics, New York.

36. Thomas, M. (2011). Fifty Key Thinkers on Language and Linguistics. Routledge: London.

37. Young, S. et al, (2002). The HTK book for version 3.2, Cambridge University.

الفصل الثَّالث

التَّحليل الصَّرفيّ الآليّ للمُفرَدات العربيَّة

د. مُحمَّد عَطيَّة

١- الكلمة كوحدة أساسية في معالجة النص العربي.

٢- أصناف الكلمات في النص العربي.

٣- طبيعة البناء الصرفي للمفردات العربية.

٤- الصيغة البنائية العامة للكلمة العربية.

٥- حجم حصيلة المفردات العربية.

٦- تعريف التحليل الصرفي الآلي وتطبيقاته.

٧- التركيب الصرفي معبرًا لحوسبة التحليل الصرفي.

٨- الالتباس الصرفي وأساليب إزالته.

٩- بناء الموارِد اللغوية اللازمة لفك الالتباس الصرفي.

١٠- أدوات وبرمجيات ومعايير.

١ - الكلمة كوحدة أساسية في معالجة النص العربي

تحفل لغتنا الجميلة بالعديد من الاستخدامات المجازية لمفردة «كَلِمة» التي من أمثلة دِلالاتها ما يلي:

١ , ١ - مَقالة؛ مثل قوله تعالى ﴿حَتَّى إِذَا جَاءَ أَحَدَهُمُ الْمَوْتُ قَالَ رَبِّ ارْجِعُونِ (٩٩) لَعَلِّي أَعْمَلُ صَالِحًا فِيمَا تَرَكْتُ كَلَّا إِنَّهَا كَلِمَةٌ هُوَ قَائِلُهَا...﴾ [المؤمنون ٩٩-١٠٠].

١ , ٢ - قرار من إنسان، أو قَدَر من الله؛ مثل قوله تعالى ﴿... ولولا كلمةٌ سبقَتْ من ربك لَقُضِيَ بينهم ... ﴾ [يونس ١٩].

١ , ٣- آية ومعجزة؛ مثل قوله تعالى ﴿... إنما المسيحُ عيسى ابنُ مَرْيَمَ رسولُ الله وكلمتُهُ ألقاها إلى مَرْيَمَ وروحٌ مِنْهُ ... ﴾ [النساء ١٧١].

... إلى آخر ما هنالك من معانٍ ودلالات عديدة.

ولكننا هنا في إطار العلوم اللسانية ومعالجة اللغة نختار[١] التعريف التقني الآتي المحدد لتلك المفردة (الكلمة هي أصغر وحدة كلامية طليقة يمكن عند كتابتها أو التلفظ بها بشكل منفرد أن تنقل للقارئ أو السامع محتوًى دِلاليًّا)، وقد أوجز كبار اللغويين العرب مثل «ابن هشام» هذا المفهوم بقولهم (الكلمةُ قولٌ مُفْرَد) وهذا بالطبع عندما نضع في اعتبارنا أن «القول» لديهم اصطلاحًا يُشْتَرَط فيه أن يحمل معنى أو دِلالة يتلقاها السامع أو القارئ.

وبالنسبة للمشتغلين بمعالجة اللغة حاسوبيًّا على المستوى الصرفي أو النحوي أو الدلالي أو ما بعد ذلك من مستويات وتطبيقات لتلك المعالجات، فإن الكلمة لديهم كذلك هي وحدة أساسية للتعامل مع النص. ولأن كل تلك المعالجات على المستوى الحاسوبي مثقلة بمشكلة الالتباس (أي احتمال أكثر من تأويل) فإن الحاسوبيين لا بُدَّ لهم من نقطة بداية غير ملتبسة يستطيعون الانطلاق منها في عملهم، ويمكننا أن نطلق على نقطة البداية هذه اسم «الكلمة الرقمية الخام» (crude digital word) ولكي نتفهم

١ - هناك جدل لغوي وفلسفي مطول بين القدماء والمعاصرين حول مفهوم وتعريف الكلمة، وربما تطلبت الإحاطة بهذا الجدل كتابًا منفصلاً، ولذا فقد أوردنا فقط أحد التعريفات العملية المقبولة على نطاق واسع.

ماهية ذلك نحتاج أولاً لإلقاء نظرة على الهيئة الرقمية التي يبدو عليها النص الخام المراد تحليله لغويًّا.

يبدو «سَيَّال النص» (text stream) الرقمي الخام عند تداوله على الحاسبات الرقمية عبارة عن سلسلة من الشفرات «الأكواد» (codes) يدل كلٌّ منها على رمزٍ معيَّنٍ واحد فقط سواء كان حرفًا هجائيًّا أم علامة ضبط صوتي أو رقمًا أو علامة ترقيم أو فراغًا ... إلى آخر ما هنالك من رموز. ويستخلص الحاسوبيون الكلمات العربية الرقمية الخام من سيال النص الرقمي الخام بالتعريف الإجرائي غير الملتبس التالي:

الكلمة العربية الرقمية الخام هي تسلسل من الشفرات يَرْمُز كل منها إما إلى حرف هجائي عربي أو علامة ضبط صوتي عربية – ولا شيء خلاف ذلك – على أن يُحَدَّ هذا التسلسل من كلا طرفيه شفرةٌ لا ترمز إلى حرف هجائي عربي ولا إلى علامة ضبط صوتي عربي.

تُمَثِّل هذه الفقرة القصيرة (الواردة في هذا الشكل رقم ١ في الفصل رقم ٣) للتعريف الإجرائي لما اصطلحنا على تسميته «الكلمة الرقمية العربية الخام»، حيث مُيِّزَت كلُّ كلمة بخلفية رمادية.

الشَّكل ٣-١ : استخلاص الكلمات العربية الرقمية الخام من النص العربي الرقمي الخام.

والشكل رقم (٣-١) أعلاه يعرض مثالاً على استخلاص الكلمات العربية الرقمية الخام من النص العربي الرقمي الخام.

تصير هذه الكلمات العربية الرقمية الخام بعد استخلاصها جاهزة لمختلَف أنواع المعالجة الحاسوبية، والتي غالبًا ما يكون أولها التحليل الصرفي الآلي وهو بدوره يكشف ما إذا كانت هذه الكلمات الرقمية الخام لا تؤدي إلى كلماتٍ غير عربية؛ سواء أكانت كلمات أجنبية مكتوبة بالحروف العربية (Transliterated words) أم كانت مجرد كلمات خاطئة غير ذات معنى، أو كانت تؤدي إلى كلمات عربية سليمة بأصنافها المختلفة وهو ما يُسْلمنا إلى القسم التالي الذي يعرض لتصنيف الكلمات العربية.

٢- أصناف الكلمات في النص العربي

عند عرض الكلمة العربية الرقمية الخام على «مُحَلِّل صرفيّ آليّ» مُعْتَبَر، فإنه من المفترض أن يستطيع هذا المحلل التعرف على تصنيف الكلمة (أو الكلمات) العربية الحقيقية التي تناظرها هذه الكلمة الرقمية الخام.

فمن الوارد ألا يتوصل المحلل الصرفي إلى أية كلمة عربية حقيقية تناظر الكلمة العربية الرقمية الخام التي قد تحتمل عندئذٍ أحد أمرين؛ إما أنها كلمة غير عربية (أجنبية) مكتوبة بحروف عربية (Transliterated) كالأمثلة التي تَرِدُ في الحالة ٢,١ بالشكل رقم (٣ - ٢) في ختام هذا القِسْم، أو أنها كلمة خاطئة كالأمثلة التي تَرِدُ في الحالة ٢,٢ في نفس الشكل. وربما يكون الخطأ في الحالة رقم ٢,٢ خطأً كليًّا بحيث تكون الكلمة الرقمية الخام عبارة عن سلسلة حروف عشوائية لا معنى لها، أو أن منشأ الخطأ خلط من الكاتب (أو من الناسخ) بين حروف الهجاء أو إسقاط بعضها أو التصاق كلمتين ... إلخ.

٢,١- تمييز الكلمات الأجنبية المكتوبة بحروف عربية

وقبل استكمال استعراض الحالات الأخرى التي يجملها الشكل رقم (٣ -٣)، فإن موضوع التمييز بين كلتا الحالتين السالف ذكرهما يستحق وقفة متأنية لأنه ليس بالأمر التافه وإن بدا كذلك لأول وهلة. فالبعض بالفعل يقارب هذه المسألة عبر إنشاء قوائم مطوَّلة بالكلمات الأجنبية التي تُكْتَب بحروف عربية، وعلى الرغم من البساطة الظاهرية لهذه المقاربة إلا أنها ملغمة بالمشكلات الخمس الآتية:

- يتغير تواتر الكلمات الأجنبية التي تكتب بحروف عربية - وهي في الغالب الأعم أسماء - باستمرار مع الزمن، فإذا تأملنا ما يرد من هذه الأسماء في العلوم المعاصرة المختلفة أو في الأخبار - على وجه الخصوص - فإننا نجد باستمرار كلمات تَجِدُّ وأخرى تختفي وبمعدل عالٍ يزيد من تسارعه حاليًّا الانفتاح الهائل في الاتصالات والعولمة، وهذا كله يجعل من تحديث مثل تلك الجداول - ناهيك عن إنشائها في الأصل - أمرًا عالي التكلفة.
- منشأ هذه الكلمات الأجنبية التي قد نصادفها مكتوبةً بالحروف العربية ضمن نصٍّ عربيٍّ قد يكون- حسب تعريفها - هو أية لغة من لغات البشر خلاف العربية بالطبع؛ وعلى سبيل المثال:

«شِنْغِهَايْ»، «بِيُونْج يَانْجْ»، «إِدْوَارْدُو»، «جُوَادَا لَاخَارَا»، «طَشْقَنْدْ»، «مَاسَاتْشُوسِتْسْ»، «بُورُونْدِي»، «مِيُونِخْ»، «بُودَابِسْتْ»، «سَاسْكَاتُونْ»، «مِسِّيسَاجَا»، «فُوكُوشِيمَا»، «كُوَالَالَمْبُورْ» ... إلخ.

وهو ما يجعل أية محاولة لوضع نماذج -ولو تقريبية- تترصد أو تتحسب لمثل هذه الكلمات صوتيًّا أو صرفيًّا -أو ما سوى ذلك- أمرًا متعذرًا.

- يزيد من صعوبة الحصر أن الهجاء الذي تكتب به كل كلمة من هذا النوع يفتقد بالطبع إلى مرجعية عربية وبالتالي إلى أية مرجعية، وعلى ذلك فهو أمر اجتهادي قد يتغير من مترجم لآخر ومن كاتب لآخر؛ ومن أمثلة ذلك الهجاءات المتداولة للكلمات الآتية:

Birmingham: «بِيرْمِينْجِهَامْ»، «بِرْمِنْجِهَامْ»، «بِرْمِينْجِهَامْ»

Sigrid Hunke: «سِيجْرِيدْ هُونْكَة»، «سِيغْرِيدْ هُونْكَا»، «سِجْرِيدْ هُونْكِهْ»، «زِيغْرِيدْ هُونْكَهْ» .... إلخ

Dayton: «دَايْتُونْ»، «دَاتُونْ»، «دِيتُونْ»

(England Englitera): «إِنْجِلْتِرَا»، «إِنْكِلْتِرَا»، «إِنْغِلْتِرَة» ... إلخ

Minneapolis: «مِينْيَابُولِيسْ»، «مِينْيَابُولِسْ»، «مِنْيَابُولِيسْ»

- كما يرتفع التحدي أكثر وأكثر أمام خيار حصر هذه الكلمات في جداول بسبب دخول اللواصق العربية على الكلمات الأجنبية المكتوبة بالحروف العربية، بل ويمكن لهذا الدخول أن يغير من هجاء هذه الكلمات؛ ومن أمثلة ذلك:

(أَكَادِيمْيَا): «أَكَادِيمِيّ»، «اَلْأَكَادِيمِيُّونَ»، «لِلْأَكَادِيمِيَّةِ» ... إلخ

(فَرَنْسَا): «الْفَرَنْسِيُّونَ»، «اَلْفَرَنْسَاوِيّ»، ... إلخ

(أَرْمِينْيَا): «الْأَرْمَنْ»، «أَرْمَنِّيّ» ... إلخ

- ويزيد الطين بِلَّةً أن كثيرًا من الكلمات الأجنبية التي تكتب بحروف عربية قد تشترك هجائيًّا مع كلمات عربية؛ مثل الأسماء الأجنبية («مِيرْ»، «مِيتْ»، «تِيمْ»،

«دُونْ»، ...) ومشتركاتها الهجائية العربية («مير»، «ميت»، «تيم»، «دون» ... إلخ).

وبسبب هذه المشكلات مجتمعةً، فإن الأنظمة الأكثر نضجًا لمعالجة اللغة العربية حاسوبيًّا تتَّبع مقاربات أخرى أكثر احترافًا في هذا الصدد يمكن تصنيفها تحت القسمين التاليين:

القسم الأول يؤجِّل التعامل مع هذه المسألة بعد التحليل الصرفي إلى حين إجراء تحليلات لغوية من طبقات أعلى وذلك مثل «تَعيين أقسام الكلام» ((Part-of-Speech Tagging و/ أو الدِّلالة و/ أو النحو، حيث تعتبَر حينئذٍ المسألة جزءًا من مسألة تمييز «أسماء الأعلام» (Proper Names) عن سواها من كلمات النص والتي توظَّف لها عادةً أساليب تعتمد على بعض القرائن المرجِّحة من مخرجات تلك الطبقات اللغوية اللاحقة إلى جانب النماذج الإحصائية الرياضية.

القسم الثاني يُخْرِج هذه المسألة من إطار طبقة التحليل الصرفي، ويصدِّرها ببساطة إلى مستوى التطبيق القائم على هذه المعالجة اللغوية. والمنطق وراء ذلك هو أن الاحتمال كبير أن يكون الحل ضمن التطبيق أكثر يُسْرًا وأقل التباسًا وذلك بما يتوافر عندئذٍ من معلومات إضافية – أو حتى قيود إضافية – كما أن التطبيق ربما يتطلب مجرد تمييز جزئيٍّ فقط أو حتى لا يتطلب من الأصل تمييز الكلمات الأجنبية المكتوبة بحروف عربية. وعلى سبيل المثال؛ فإن تطبيقات الربط المعجمي مثلاً قد لا تهتم كثيرًا بهذا التمييز، في حين أن تطبيقات البحث النصي أو استرجاع المعلومات قد يكون الوعاء النصي الذي تبحث فيه مقيِّدًا لاحتمالات تأويل الكلمات سواء العربية أم الأجنبية والتمييز بينهما فضلاً عن أن هذا النوع من التطبيقات عمومًا غير مرهف الحساسية للأخطاء في هذا الصدد، أما تطبيقات التشكيل الصوتي الآلي فهَمُّها الرئيسي هو دقة استنباط الضبط الصوتي لكل الكلمات المكتوبة بالنص العربي سواء أكانت عربية أم أجنبية ولذلك فقد يكون من الأفضل أن تحتفظ بكل التأويلات الممكنة وفي نهاية المطاف تفاضل بينها احتماليًّا وفق نماذج إحصائية وربما مع قيود فونولوجية تضمن موافقة الضبط الصوتي المرجَّح للقوانين الصوتية العربية [١].

١– القسم السادس من هذا الفصل مخصص لتطبيقات التحليل الصرفي الآلي للنص العربي.

٢ , ٢- تصنيف الكلمات العربية

وبالعودة الآن إلى استعراض بقية الحالات التي يمكن أن ينتجها محلل صرفي معتبَر عند تشغيله على كلمة عربية رقمية خام، فإنه يمكن أن يعثر على تركيب صرفي (أو تراكيب صرفية) تناظر كلمة (أو كلمات) عربية تقابل هذه الكلمة الخام وهو المسار الرئيسي الأول حسب الشكل رقم (٣-٢) أدناه.

يقودنا هذا المسار بدوره إلى أحد مسارين؛ إما إلى كلمة عربية مصرَّفة أو كلمة عربية غير مصرَّفة، ونبدأ بمسار الكلمة العربية المصرفة وهي الحالة رقم ١ , ١ في الشكل رقم (٣-٢).

فاللغة العربية تؤول الغالبية الكاسحة من كلماتها إلى مواد لغوية أولية يُطْلَق عليها «جُذور» وكل جذر منها هو عبارة عن؛ إما ثلاثة حروف - وهو الأكثر - ويسمى عندئذٍ «جِذرًا ثلاثيًّا» مثل مادة (ن، ص، ر)، أو أربعة حروف - وهو الأقل - ويسمى عندئذٍ «جِذرًا رباعيًّا» مثل مادة (ع، ر، ق، ل). وتتولد الكلمة من صب إحدى هذه المواد في أحد القوالب (الصِّيَغ) الصرفية القياسية، والصيغة الصرفية هي كلمة بضبط صوتي كامل إلا أن بها ثلاثة حروف (إذا كانت صيغةً للجذور الثلاثية) أو أربعة حروف (إذا كانت صيغةً للجذور الرباعية) متغيرة قابلة لاستبدالها بحروف الجذر عند صبه في هذه الصيغة.

فصيغة «فَاعِل» على سبيل المثال هي صيغة تقبل الجذور الثلاثية وتكون حروف الفاء والعين واللام في الصيغة حروفًا عامة تقبل استبدالها بحروف الجذر الثلاثي على الترتيب؛ فإذا صببنا على سبيل المثال جذر «ن ص ر » في هذه الصيغة (ف=ن،ع=ص، ل=ر) تتولد كلمة «نَاصِر»، وبنفس الأسلوب إذا صببنا مادة «ع ر ق ل» في صيغة للجذور الرباعية مثل «فَعَالِيل» - حيث يحل الحرف الرابع من الجذر محل اللام الثانية من الصيغة الصرفية - تتولد كلمة «عَرَاقِيل» وكذلك إذا صُبَّت نفس المادة في الصيغة الفعلية «فَعْلَلَ» تتولد كلمة «عَرْقَلَ»... وهكذا تسير عملية اشتقاق الغالبية العظمى من الكلمات العربية التي هي كلمات عربية منتظمة التصريف.

إلى جانب ذلك يوجد عدد محدود جدًّا (بضع مئات قليلة) من الكلمات العربية غير منتظمة التصريف تَوَلَّدَ كُلٌّ منها نتيجة صب أحد الجذور في قالب صرفي غير قياسي (أي أنه قالب غير صالح لتوليد كلمات أخرى غير هذه الكلمة بالتحديد) أو ربما كان تصريفها قياسيًّا في البداية ثم وقع شذوذ عنه ثم شاع الشذوذ - لسهولة نطقه مثلاً - بينما

خَفُتَ القياس. ومن أمثلة ذلك الكلمة الفعلية «اتَّخَذَ» من مادة (أ، خ، ذ) وكان المفترض أن يؤدي تصريفها القياسي إلى كلمة «ائْتَخَذَ» على صيغة (افْتَعَلَ) القياسية وذلك مثل نظيراتها «ائْتَمَنَ» من (أ، م، ن) وكذلك «ائْتَمَـــرَ» مـــن (أ، م، ر)... إلخ.

ومحصلةُ ذلك أن مسار الكلمة العربية المصرفة [1] يتفرع بدوره إلى أحد مسارين؛ إما إلى مسار الكلمة العربية منتظمة التصريف (المرقَّم ١ , ١ , ١ في الشكل رقم ٣-٢) أو إلى مسار الكلمة العربية غير منتظمة التصريف (والمرقَّم ١ , ١ , ٢ في نفس الشكل).

أما مسار الكلمات العربية غير المصرفة - ورقْمُه ١ , ٢ في الشكل رقم (٣-٢) - فهو يختص بنوعين من الكلمات لا تتولد نتيجة عملية صرفية كالتي عرضناها للتو، ولكنها دخلت اللغة كما هي على حالها [2] من حيث ثبات الهجاء والضبط الصوتي (ربما باستثناء تَغيُّر الحركة الصوتية الإعرابية مع تغير المحل الإعرابي لما هو مُعْرَب من هذه الكلمات). ويقودنا هذا المسار بدوره إلى أحد مسارين؛ إما إلى كلمة جامدة وهي الحالة رقم ١ , ٢ , ١ في الشكل رقم (٣-٢) أدناه، أو إلى كلمة معَرَّبة وهي الحالة رقم ١ , ٢ , ٢ في نفس الشكل.

أمَّا الكلمات الجامدة فهي بضع مئات قليلة من الكلمات العربية لكنها عالية التكرار في النص العربي، وأغلبها أدوات كحروف جر، وحروف ناسخة، وأسماء إشارة، وأسماء موصولة، وحروف نداء، وضمائر منفصلة ... إلخ؛ وذلك مثل «مِنْ»، «إِنَّ»، «ذَا»، «الَّتِي»، «يَا»، «نَحْنُ»، ... إلخ.

وأمَّا الكلمات المعَرَّبة فهي أيضًا بضع مئات قليلة من الكلمات التي دخلت اللغة العربية مـــن لـــغات الحضارات المجـــاورة (كالفــــــارسية والحبشـــية والســـريانية والرومـــانية....) واستقرت فيها وربما عَدَّل العرب من أصواتها لتلائم ذائقتهـــم اللغوية، ومما يدل على استقرار هذه الألفاظ في اللغة العربية أن القرآن الكريم وقد نزل بلسان عربيٍّ مبين قد أورد بعضًا منها. ومن أمثلة تلك الكلمات المعرَّبة؛ «إِسْتَبْرَق» (وهي من «إِسْتَبْرَه» الفارسية)، «آمِينْ» (وأصلها حبشي)، «مُوسَى، هَارُونْ، يُوسُف، مَرْيَم ... إلخ» (وهي أسماء عبرانية)، ... إلخ.

١- يحتوي القسمان الثالث والرابع من هذا الفصل على المزيد حول طبيعة البناء الصرفي للمفردات العربية.

٢- مثل كل أنواع الكلمات العربية، فإن هذه الكلمات غير المصرفة تقبل إضافة اللواصق إليها (من سوابق ولواحق) تحت شروط معينة، ويحتوي القسمان الثالث والرابع من هذا الفصل على المزيد حول هذا الأمر.

ومن الهام في هذا المقام الإشارة إلى أن اللغة العربية تعتمد أسلوبًا آخَرَ هامًا للتعريب أكثر مرونة من هذه الفئة المحدودة من الكلمات المعَرَّبة، ويعتمد هـذا الأسلوب على اختيار الحروف الأربعة الأكثر تمييزًا لِلَّفْظِ الأجنبي وجعلها مادةً لجذر رباعي يمكن تصريفه كما سبق شرحُه عالِيَهُ في الحالة رقم ١ , ١ , ١ فعلى سبيل المثال؛ الجذر الرباعي (ت، ل، ف، ز) مستخلَصٌ من الكلمة الفرنسية «تِلِفِزْيُون» ويُشْتَق منه «تَلْفَز»، «تَلْفَزَة»، «تِلْفَاز»، «مَتَلْفَز» ... إلخ، كما أن الجذر الرباعي (هـ، ن، د، س) مستخلَصٌ من الكلمة الفارسية «إِنْدَازَة» ويُشْتَق منه «هَنْدَسَ»، «هَنْدَسَة»، «مُهَنْدِس»... إلخ.[1]

كلمة عربية رقمية خام

١- عربية سليمة

١ , ١- مصرَّفة

١ , ١ , ١- منتظِمة

أمثلة: لِلْبَاحِثِينَ، فَالْبُلْدَان، يَقْرَؤُونَهُ ...

١ , ١ , ٢- غير منتظِمة (استثناء صرفيّ)

أمثلة: الله، تَقْوَى، تُهم ...

١ , ٢- غير مصرَّفة

١ , ٢ , ١- جامدة

أمثلة: هُوَ، الَّذِي، مِنْ، مَعَهُمْ ...

١ , ٢ , ٢- معَرَّبة

أمثلة: آمِينْ، إِبْرَاهِيمْ، إِسْتَبْرَق، جَهَنَّم ...

٢- غير عربية

٢ , ١- أجنبية مكتوبة بحروف عربية

أمثلة: وَاشِنْطُنْ، تُرْكُمَانِسْتَانْ، بَرْلَمَانِيَّة، اَلتِّكْنُوقْرَاطِيَّة ...

٢ , ٢- خاطئة

أمثلة: لنبنمسبمن، ألظباط، بكقاجيكت، كالمبينبالشكل...

الشَّكل ٣-٢: تصنيف ما يمكن أن تؤول إليه أية كلمة عربية رقمية خام بعد تحليلها صرفيًّا.

١- من الأمثلة الأكثر حداثة أيضًا؛ الجذر الرباعي (أ، ت، م، ت) لتعريب كلمة «automation» الدالة على التشغيل الآلي أو التلقائي لمهمةٍ ما، وكذلك الجذر الرباعي (ق، و، ق، ل) من اسم شركة «Google» ليدل على البحث في «الشبكة العنكبوتية Internt».

٣- طبيعة البناء الصرفي للمفردات العربية

يحتوي جسم الكلمة العربية دائمًا على جزء أساسي يُطْلَق عليه اصطلاحًا اسم «الجِذْع» (Stem) وهو الجزء الذي جرى للتو عرض أصنافه وكيفية نشوء كل منها في القسم السابق. لكن الجذع لا يكوِّن بمفرده الكلمة العربية في حالتها العامة، بل إنه يمكن أن يلتصق بجزء آخَر يتقدم عليه ويُطْلَق عليه اصطلاحًا اسم «السابقة» ((prefix، كما يمكن أن يلتصق بالجذع جزء ثالث يَتْبعُهُ ويُطْلَق عليه اصطلاحًا اسم «اللاحقة» (suffix). ويُصْطَلَح على تسمية السوابق واللواحق معًا «باللواصق» (affixes). وبطبيعة الحال فإن اللواصق تضاف إلى الجذع لإنتاج كلمة توائم بنيتُها الوظيفية سياقَها النحويَّ وكي تكتمل دِلالتها في السياق المعنوي للكلام.

فعلى سبيل المثال إذا كان لدينا جذع الكلمة «رِهَان» – من مادة «ر هـ ن» المصبوبة في الصيغة الصرفية «فِعَال» – فإن إضافة السابقة «الـ» إليها ينتج كلمة «الرِّهَان»، فإذا أضفنا إليها اللاحقة «ات» صارت «الرِّهَانَات». وفي حـــين أن الســـابقة في هذا المثـــال قد أضافـــت سِمَة التعريف فإن اللاحقة قد أضافت سِمَتَي الجمع والتأنيث، فإذا كانت هذه الكلمة قد وردت – على سبيل المثال – في عبارة (... مع أن هذه الرِّهَانَاتِ الكبيرةَ كانت محل تساؤل منذ البداية ...) فإن إفادة السابقة في هذا السياق هو التعريف بغرض التخصيص؛ بما يعني هذه الرهانات بالتحديد وليس أية رهـــانـــات أخـــرى، كمـــا أن اللاحقة أفادت الجمع ولأن المفردة تدل على غير عاقل فإنها قد استخدمت صيغـــة جمع المؤنث، ونلاحظ كذلك ما تقتضيه سلامة التركيب النحوي من موافقة الصفة «الكبيرة» لموصوفها «الرهانات» من حيث التعريف والنوع (التأنيث في هذه الحالة) بينـما احتفظـــت الصفة بالإفراد لكون موصوفها غير عاقل ويماثل الصفـــة في ذلك كلٌّ من اسم الإشـــارة «هـــذه» الذي يسبق الكلمـــة والفعل الذي يعود عليها «كانـــت»[١].

ومن أجل حوسبة البِنْية الصرفية للكلمة العربية، علينا أن ننتبه إلى أربع مسائل أساسية وهي مسألة اللواصق المركَّبة، ومسألة ضمان عدم التنافر بين أي زوجين من

١– لاحظ التغير في سمة العدد بالنسبة للصفة واسم الإشارة والفعل عندما يكون الموصوف عاقلاً؛ كأن تصير العبارة كما يلي: (... مع أن هؤلاء السيدات الكبيرات كُنَّ محل ترحيب منذ البداية ...).

الأجزاء الثلاثة لأية كلمة سليمة، ومسألة التأثير المتبادل في بنية الأجزاء الثلاثة عند التحامها في كلمة واحدة، ثم مسألة التوافق مع الفونولوجيا العربية عبر الإعلال والإبدال والحذف.

٣,١- اللواصق البسيطة واللواصق المركبة

فمن حيث المسألة الأولى، فإن السابقة قد تكون بسيطة ذات عنصر واحد؛ وذلك مثل («الـ» التعريف، «بـ» الجر، «يـ» المضارَعة، ... إلخ) كما أنها قد تكون مركَّبة من أكثر من عنصر واحد؛ وذلك مثل («بِالـ»، «فَسَيـ»، ... إلخ) وذلك كما في «بِالأدوات» و«فَسَيَعلمون» على الترتيب. وكذلك هو الحال بالنسبة لِلَّاحقة التي قد تكون بسيطة ذات عنصر واحد؛ وذلك مثل («ـة»، «ـوا»، «ـهم»، ... إلخ) كما أنها قد تكون مركَّبة؛ وذلك مثل («ـتَانِ»، «ـوهُم»، ... إلخ) وذلك كما في «كاتبتَانِ» و «لِيُساعدوهُمْ» على الترتيب.

وفي حين أن المقاربة الأصعب تسلك سبيل توصيف اللواصق البسيطة فقط حاسوبيًّا لبرمجيات التحليل الصرفي ومن ثَمَّ إلقاء عبء رد اللواصق المركَّبة إلى مكوناتها البسيطة على عاتق المعالجة الحاسوبية لتلك البرمجيات، فإن المقاربة الأيسر والأكثر واقعية تقوم بحصر كل السوابق المركَّبة وكل اللواحق المركَّبة الممكنة في اللغة العربية - وهو أمر ممكن حيث أن عدد كل منها بضع مئات محدودة - ثم يَجري بعد ذلك التوصيف الحاسوبي في برمجيات التحليل الصرفي لكلٍّ من تلك اللواصق من حيث الهجاء والضبط الصوتي والسِّمَات والفِعْل البنيوي وهذان العنصران الأخيران في التوصيف هما اللذان يسلمانِنا للمسألتين التاليتين.

٣,٢- شروط التحام اللواصق بجذع الكلمة

أما المسألة الثانية فهي أن التحام اللواصق بجذع الكلمة لا بُدَّ أن يحقق شروط عدم التنافر بين أي جزأين من الأجزاء الثلاثة الأساسية للكلمة (السابقة، الجذع، اللاحقة)، فيمكن - على سبيل المثال - أن تلتحم سابقة «الـ» التعريف مع جذع الكلمة الاسميّ «صَالِـح» مع اللاحقة «ة» لتكوين الكلمة «الصَّالِحَة» لأن هذه الأجزاء الثلاثة لا يتنافر فيها أي زوجين، كما يمكن أيضًا على سبيل المثال

أن تلتحم سابقة «يـ» المضارَعة مع جذع الكلمة الفعلي «تَنَافَس» مع اللاحقة «ـون» لإسناد الفعل المضارع إلى جمع المذكر ورفعه بثبوت النون لتكوين الكلمة «يَتَنَافَسُون»، وذلك لأنه في كــلا المثــالين لا يتنافر أي جزأين من الأجزاء الثلاثة للكلمة. بينما لا يمكن مثلاً أن تلتحم سابقة «يـ» المضارعة مع جذع الكلمة الاسمي «مَصَابِيح»، ولا يمكن كذلك أن تجتمع في كلمة واحدة سابقة «سَيـ» للتسويف والمضارَعة والتي تقتضي لزوم الفعل حالة الرفع - مع لاحقة «ـوا» التي تشير إلى حالة نصب أو جزم الفعل المضارع (أو الأمر) المسند لجمع المذكر.

وللتأكد من تَحَقُّق شروط عدم التنافر هذه يمكن اتباع أسلوب رياضي يسهل تنفيذه حاسوبيًّا ومؤداه أن يجري تعريف فئة محدودة من السمات الأساسية - يكون عددها في حدود عشرات قليلة - التي تغطي كل السمات التي يمكن أن تتخذها أيٌّ من السوابق { اسمية، محايدة، فعلية، ... إلخ}، وكذلك يجري تعريف فئة أخرى للسمات الأساسية لِلَّواحق، وبناءً على ذلك يجري إعداد جدول تبين خلاياه الحالة بين كل عنصرين من هاتين الفئتين من حيث التنافر (تحتوي الخلية عندئذٍ على صفر) أو التواؤم (وتحتوي عندئذٍ على واحد).

وأثناءَ إعداد برمجيات التحليل الصرفي يُضَمَّن التوصيفُ الحاسوبي لكل لاصقة - سواء كانت سابقة أم لاحقة - شفراتِ (أكواد) السمات التي تنطبق عليها، وكذلك يُضَمَّن التوصيف الحاسوبي لكل قالب صرفي (نِيابةً عن الجذع) السماتِ غير المقبولة في اللواصق التي تلتحم معه (وأحيانًا سمةً يتحتم وجودها في اللاصقة التي تلتحم معه).

وتأسيسًا على كل ما سبق يمكن أن تتحقق برمجيات التحليل الصرفي في «وقت التشغيل» ((runtime من تواؤم إحدى السوابق مع إحدى اللواحق ببساطة عن طريق التأكد أنه لا توجد أية سمة لِلَّاصقة تتنافر مع أية سمة لِلَّاحقة وذلك بمراجعة الخلايا المناسبة في الجدول المشار إليه عالِيَه، ويمكن بعد ذلك التأكد من أن توصيف سمات كل من السابقة واللاحقة لا يحتوي على أية سمة غير مرغوب فيها من قبل القالب الصرفي[1].

---

١ - وكذلك أن توصيف سمات كل من السابقة واللاحقة يحتوي على أية سمة يتحتم وجودها في القالب الصرفي.

٣,٣- التأثير البنيوي المتبادل بين اللواصق والجذع

وتتعلق المسألة الثالثة بأن السابقة والجذع عند التحامهما يؤثران بشكل متبادل في بعضهما البعض من حيث الهجاء والضبط الصوتي، وكذلك هو الحال عند التحام الجذع واللاحقة معًا، فعلى سبيل المثال؛ إذا التحمت السابقة «بِال» بالجذع «اقْتِصَاد» (الناتج عن صب مادة الجذر «ق ص د» في القالب الصرفي «افْتِعَال») فإن ألف الوصل التي تتصدر الجذع تصير غير منطوقة دائمًا بسبب اتصالها بالسابقة (وهذا هو فِعْل السابقة في الجذع) ولكن ذلك يؤدي إلى تتابع ساكنَيْن (اللام في السابقة والقاف في الجذع) وهو ما يخالف القوانين الصوتية (الفونولوجية) العربية وللخروج من ذلك يُحَرَّك الساكن الأول وهو اللام بالكسر (وهذا هو فِعْل الجذع في السابقة) وتكون نتيجة هذا الالتحام هي «بِالِاقْتِصَاد». وعلى الجانب الآخر، إذا التحم الجذع «صَمَتَ» (الناتج عن صب مادة الجذر «ص م ت» في القالب الصرفي «فَعَلَ») مع اللاحقة «تُم» - وهي ضمير الرفع للمخاطَبِين - فإن هذه اللاحقة تقتضي تسكين آخر حروف الفعل الماضي وهي التاء (وهذا هو فِعل اللاحقة في الجذع) فيصير لدينا حرفان متشابهان أولهما ساكن وهو ما تقتضي القوانين الصوتية العربية إدغامهما معًا في حرف مشدد (وهذا هو فِعل الجذع في اللاحقة) فتكون نتيجة هذا الالتحام هي «صَمَتُّم».

وللتعامل حاسوبيًّا مع هذه المسألة فإنه يجري تعريف فئة محدودة العدد من التأثيرات التي يمثل كلٌّ منها فِعْلاً متبادَلاً بين بنية السابقة والجذع، كما يجري كذلك تعريف فئة أخرى محدودة العدد من التأثيرات التي يمثل كلٌّ منها فِعْلاً متبادلاً بين بنية الجذع واللاحقة. ويكون كلٌّ من هذه التأثيرات في كلتا الفئتين عبارة عن إجراء حاسوبي مكون من خطوات مشروطة لتعديل بِنْيَتَي الجزأين الملتحمين حسب ما يصادفه الإجراء من حروف وحركات في السابقة والصيغة الصرفية الملتحمتين أو الصيغة الصرفية واللاحقة الملتحمتين. ويجري بطبيعة الحال تضمين رمز لهذا الإجراء في التوصيف الحاسوبي لكل لاصقة أثناء إعداد برمجيات التحليل الصرفي، كما يُضَمَّن التوصيفُ الحاسوبيُّ لكل صيغة صرفية «رايةً» ((flag وهي عبارة عن رمز رقمي ينتمي لفئة محدودة العدد) قد تدخل قيمتها ضمن شروط تنفيذ بعض خطواتِ التأثيراتِ السابقِ ذِكرُها.

٣, ٤- الإعلال والإبدال والحذف

أما المسألة الرابعة والأخيرة فهي متعلقة بتكوين جذع الكلمة عبر صب مادة الجذر في الصيغة الصرفية (وتسمى أيضًا «القالب الصرفي» وكذلك «الميزان الصرفي»)، وهو ما مررنا عليه سريعًا في القسم السابق ونود هنا أن نلقي الضوء على مزيد من التفاصيل الهامة بخصوصه. فبالنسبة لعدد الصيغ الصرفية الأساسية في اللغة العربية بما فيها الصيغ الاسمية والفعلية وبما فيها صيغ الجذور الثلاثية والرباعية فإن عددها جميعًا قد يربو قليلاً على المئة صيغة، هذا بافتراض أن كل حروف الصيغة الصرفية المتغيرة يجري التعويض عنها بشكل مباشر بحروف مادة الجذر وهي الحالة التي عرضناها ومَثَّلْنا لها في القسم السابق، ولكن الأمر في الواقع أكثر تعقيدًا من ذلك.

ولتفهم ذلك ينبغي علينا أن نتذكر أن الصرف طبقة لغوية تالية للطبقة الصوتية (الفونولوجية) وهذا يعني أن البناء الصرفي العربي محكوم بالقوانين الفونولوجية العربية، وفي حين أن هذه القوانين لا تمنع التعويض المباشر في الحروف المتغيرة للصيغة الصرفية حين تكون الحروف في مواضع «صِحَّة/ صحيحة»؛ وذلك مثل صب مادة (ك، ث، ر) في القالب الصرفي «تَفَاعُل» حيث نحصل بالتعويض المباشر (ف=ك، ع=ث، ل=ر) على جذع الكلمة «تَكَاثُر»، ولكن بعض الحروف قد تقع في مواضع «عِلَّة» أو مواضع «إبدال» صوتية وهو ما يجعل التعويض المباشر في هذه المواضع مخالفًا للقوانين الفونولوجية العربية فيتوجب عندئذٍ من أجل موافقتِها تبديلُ الحرف عند أحد هذه المواضع بحرف آخر أو حتى حذفُه أحيانًا.

ونضرب لذلك بعض الأمثلة؛ فلو عوضنا بشكل مباشر عند صب مادة (س، ق، ي) في صيغة «اسْتِفْعَال» لحصلنا على «اسْتِسْقَاي» وهو مستهجَن فونولوجيًّا في العربية لتطرف الياء بعد ألف ممدودة مما يقتضي تحولها إلى همزة لتصير «اسْتِسْقَاء». ولو عوضنا بشكل مباشر عند صب مادة (و، ل، د) في صيغة «مِفْعَال» لحصلنا على «مِوْلاد» وهو غير مقبول فونولوجيًّا في العربية لأن الواو الساكنة تلي حركة الكسرة ولابد لها من أن تتحرك على نفس جنس هذه الحركة فتصير ياءً ممدوة فنحصل على «مِيلاد». ولو عوضنا بشكل مباشر عند صب مادة (ض، ر، ب) في صيغة «افْتِعَال» لحَصُلْنا على «اضْتِرَاب» وهو متعذر النطق لصعوبة الانتقال من مخرج الضاد - وهو حرف مفخَّم

– بعد السكون عليه إلى المخرج البعيد عنه للتاء – وهو حرف مرقَّق – فيتوجب لذلك تحريك حرف التاء إلى مفخَّمها فتصير طاءً يمكن الانتقال إليها بيُسْر بعد الضاد الساكنة فنحصل على «اضْطِرَاب». ولو عوضنا بشكل مباشر عند صب مادة)ت، ل، و) في صيغة فعل الأمر «افْعُلْ» لحصلنا على «اتْلُوْ» – هذا بالطبع إذا لم تكن هناك لاحقةٌ بعد الفعل؛ أي إذا كان الأمر مسنَدًا إلى مفرَد مخاطَب – وهو عسير النطق بسبب صعوبة المحافظة على تسكين الواو بعد حركة الضمة حيث أن الفونولوجيا حينئذ تدفع نحو تحريك الواو بالمد ولكن ذلك يطمس علامة إعراب الفعل المجزوم، ولذلك تُحْذَف الواو كعلامة إعراب ونحصل على «اُتْلُ» كفعل أمر مجزوم وعلامة جَزْمِه حذف حرف العلة وهو الواو في مثالنا هذا.

وتستلزم تغطية حالات الإعلال والإبدال والحذف في كل الصيغ الصرفية العربية إضافة توصيف حاسوبي لِصِيَغ صرفية فرعية تصف كل منها حالة من تلك الحالات، وعلى ذلك يتجاوز العدد الإجمالي للصيغ الصرفية الألف بقليل، كما تُعَيَّن الحروف المتغيرة في القسم الهجائي من التوصيف الحاسوبي لكل من هذه الصيغ برموز تدل على ما إذا كانت صحيحة أو معتلة أو محذوفة فضلاً بالطبع عن ترتيب كلٍّ منها بالمناظَرة مع حروف الجذور. وجدير بالذكر أن التوصيف الصوتي للصيغة الصرفية هو ذاته التوصيف الصوتي لكل الجذوع التي يمكن أن تنتج عن صب مادة أي جذر فيها، كما أن محدِّدات قابلية التحام هذه الجذوع باللَّواصق – كما سبق التنويه إليه في القسم الفرعي قبل السابق – يتضمنها أيضًا التوصيف الحاسوبي للصيغة الصرفية [1].

٤– الصيغة البنائية العامة للكلمة العربية

بِناءً على كل ما عرضناه في القسمين السابقين (الثاني والثالث) من هذا الفصل، فإن الكلمة العربية تتولد عبر الخطوات الثلاث التالية:

أ. تُصَبُّ أولاً مادةُ إحدى الجذور في إحدى الصيغ الصرفية التي يسمح توصيفها الحاسوبي باستقباله (وفق القوانين الفونولوجية العربية) وذلك لإنتاج جذع

١– لمزيد من التفاصيل التقنية الدقيقة لكل ما ورد في هذا القسم يمكن الاطلاع على الفصل الخامس بالمرجع رقم [١٥] فى قائمة المراجع نهاية هذا الكتاب.

الكلمة، ونرمز لهذه العملية كما يلي:

جذع كلمة = جذر ⊙← صيغة صرفية

ب. تلتحم إحدى السوابق مع جذع الكلمة الناتج إذا تحققت شروط عدم التنافر بين السابقة والصيغة الصرفية للجذع، مع أخذ التأثير البنيوي المتبادل بين السابقة والجذع في الاعتبار، ونرمز لذلك كما يلي:

سابقة ⊕←جذع كلمة = سابقة ⊕←(جذر ⊙← صيغة صرفية)

ج. تلتحم إحدى اللواحق مع الناتج من الخطوة السابقة إذا تحققت شروط عدم التنافر بين اللاحق والصيغة الصرفية، مع أخذ التأثير البنيوي المتبادل في الاعتبار، فنحصل بذلك على الكلمة في شكلها النهائي. ويمكننا صياغة هذه الخطوات رمزيًّا كما يلي:

كلمة عربية = سابقة ⊕←(جذر ⊙← صيغة صرفية)→⊕ لاحقة

الشَّكل ٣-٣: الصيغة البنائية العامة للكلمة العربية.

وتُعَدُّ هذه الصيغةُ ركيزةً أساسية لحوسبة الصرف العربي عبر برمجيات التحليل الصرفي العربي، هذا بالطبع شريطةَ أن تستطيع بالفعل استيعابَ أية كلمة تقع تحت أيٍّ من أصناف «الكلمات العربية السليمة» المبيَّنة في الشكل رقم (٣-٣) عالِيَه. ففي حين أن القسم السابق قد أوضح بالتفصيل دقة توصيف هذه الصيغة البنائية لبنية الكلمات العربية منتظمة التصريف، فإنها أيضًا تنطبق على الكلمات العربية استثنائية التصريف حيث يكون جذع كل استثناء صرفي حالةً فريدةً لا تقبل معه الصيغة الصرفية الاستثنائية أن يُصَبَّ فيها إلا جذر وحيد ولذلك فإن توصيفها الهجائي لا يحتوي على أية حروف متغيرة.

ويتبقى الآن نوعان من الكلمات العربية غير المصرفة؛ وهما الكلمات الجامدة (وعددها مئات قليلة) والكلمات المعرَّبة (وعددها كذلك مئات قليلة)، وكلاهما بحاجة إلى أن يجد مكانًا وتفسيرًا في هذه الصيغة البنائية العامة للكلمة العربية. فعَلَى ناحية الصيغة الصرفية فمن الواضح أنه يمكن مناظرتها بسهولة – في كلا النوعين –

بحالة الاستثناء الصرفي أي أن الجذع الجامد أو المعَرَّب يلعب هو نفسه دور الصيغة الصرفية بلا أية حروف متغيرة، أما على ناحية الجذر فيبدو لأول وهلة أنه عديم القيمة لأنه لا يلعب دورًا في توليد الجذع ولا الكلمة، ولكننا نستطيع أن نعطيه دورًا قد يُفِيد في بعض التطبيقات للمعالجة الحاسوبية للغة العربية.

فمن المفيد بالنسبة للجوامد أن ننظر إلى المترابط منها وظيفيًّا كعائلة ننتخب إحداها (ويفضَّل أن يكون أبسطها تركيبًا، أو أكثرها شيوعًا) كجذر لهذه العائلة، فعلى سبيل المثال؛ الأسماء الموصولة {«مَنْ»، «مَا»، «الَّذِي»، «الَّتِي»، «الَّذِين»، «اللَّاتِي»..} يمكن أن ننظر إليها كعائلة ونضع لها جذرًا وليكن «الَّذِي»، كما يمكن أن ننظر إلى ضمائر الرفع المنفصلة في حالة الخطاب {«أَنْتَ»، «أَنْتِ»، «أَنْتُما»، «أَنْتُمْ»، «أَنْتُنَّ»} كعائلة ونضع لها جذرًا وليكن «أَنْتَ»، ... وهكذا. وفي حالة وجود جامدة غير مرتبطة بغيرها من الجوامد فإن الجذر ببساطة يكون مساويًا للجذع وكفى.

أما في حالة الكلمات المعرَّبة فتلك العائلات تضم إما الكلمة المعرَّبة بمفردها، أو الكلمة المعرَّبة بتفرعاتها الهجائية المختلفة؛ وذلك مثل {«مُوسِيقَى»، «مُوسِيقَا»، «مُوسِيقَار»..} ونضع لها جذرًا وليكن «مُوسِيقَى»، وكذلك {«مِيكَال»، «مِيكَائِيل»، «مِيخَائِيل»} ونضع لها جذرًا وليكن «مِيخَائِيل»، ... وهكذا.

ويفيد تجميع هذه الكلمات في عائلات في بعض التطبيقات الحاسوبية الهامة لمعالجة اللغة العربية؛ ومنها على سبيل المثال «البحث في النصوص العربية»، و«استرجاع المعلومات»، وكذلك في بعض نواحي التحليل الدلالي العربي.

وتأسيسًا على كل ما تقدَّمَ فإنه إذا توافر لدينا جذرٌ، وصيغةٌ صرفيةٌ، وسابقةٌ، ولاحقةٌ، فإننا نستطيع عبر الصيغة البنائية العامة للكلمة العربية أن نقوم حاسوبيًّا بتركيب الكلمة العربية المناظرة لتلك المكوِّنات (بافتراض عدم تنافرها)، مع التنويه أننا بالطبع نحتاج أيضًا لمعرفة نوع الكلمة العربية من أجل تقرير كيفية تركيب جذعها من جذرها وصيغتها الصرفية، ولذلك فإن التركيب (نوع الكلمة: السابقة، الجذر، الصيغة الصرفية، اللاحقة) هو تمثيلٌ كافٍ ومكافئٌ للكلمة العربية، مع ملاحظة أن غياب السابقة عن أية كلمة يُعْتبَر بحد ذاته سابقة (سابقة فارغة null prefix) وبالمثل فإن غياب اللاحقة في أية كلمة يُعتبَر لاحقة (لاحقة فارغة

.(null suffix

ومن الواضح في هذا التمثيل المعلوماتي أنه يكفي ذِكْر «الكود» الرقمي لكل مُكَوِّن من هذه المكوِّنات الخمس والذي يُمْكِن من خلاله الوصولُ إلى التوصيف الحاسوبي الكامل لهذا المكون في برمجيات التحليل الصرفي.

٥- حجم حصيلة المفردات العربية

ما هو حجم «حصيلة مفردات» (vocabulary) اللغة العربية؟

رَغْمَ أن هذا السؤال يبدو بريئًا بسيطًا، إلا أن صياغته تحتاج أولاً إلى المزيد من التحديد.

كما أن إجابته بعد ذلك تحتاج إلى نظر عميق وتفصيل!

فمن حيث تحديد السؤال؛ فهناك حصيلة المفردات على «القِيَاس»، وهناك حصيلة المفردات على «السماع»، وبين الاثنتين فارق في الحجم كبير.

فحصيلة المفردات العربية القياسية تشمل كل الكلمات العربية السليمة الممكن توليدُها من المكوِّنات الصرفية العربية الأساسية - أيْ؛ الجذور، والصِّيَغُ الصرفية، والسوابق، واللواحق - عَبْرَ آليات وضوابط التوليد والالتحام السالف ذكرها في القسم السابق، أما حصيلة المفردات العربية السماعية فهي تشمل كل الكلمات العربية السليمة التي استخدمها بالفعل الناطقون باللغة العربية تحدثًا وكتابةً على مر مختلَف العصور وفي مختلَف المجالات وفي مختلَف المناطق الجغرافية وعبر مختلف أساليب التحدث (سردًا، وحوارًا، وخطابة...).

أما على جانب تقدير حجم حصيلة المفردات على القياس، فإن المحاولات المعتبَرة لبناء برمجيات الصرف العربي تُنْبِئُنا أن أحجام فئات المكوِّنات الأساسية (اللَّبِنات) لبناء الكلمات العربية هي في الحدود التالية:

عدد الجذور المصرَّفة[1] ≈ ٤,٥٠٠

١- بعض المعاجم العربية المطوَّلة [٢] التي تُعْنَى بجمع كل الروايات تصل بهذا العدد إلى حوالي تسعة آلاف جِذر، ولكن

عدد الصِّيَغ الصرفية المنتظمة الأساسية	≈	١٠٠
عدد جميع الصيغ الصرفية المنتظمة[١]	≈	١,١٠٠
عدد الاستثناءات الصرفية	≈	٣٥٠
عدد الجوامد	≈	٣٠٠
عدد المعَرَّبات	≈	٤٠٠
عدد السوابق (البسيطة والمركَّبة)	≈	٣٠٠
عدد اللواحق (البسيطة والمركَّبة)	≈	٥٥٠

ومن هذه الأرقام نستطيع أن نقدر مجموع عدد الجذوع العربية الممكنِ توليدُها بحوالي النصف مليون جذع وهو حاصل ضرب عدد الصيغ الصرفية الأساسية[٢] في عدد الجذور المصرفة، ويلاحَظ بالطبع أن عدد الاستثناءات الصرفية والجوامد والمعرَّبات معًا في حدود الألف وهو عدد مهمَل بالنسبة لنصف مليون.

أما بالنسبة لِلَّواصق فإن عدد الأزواج المرتَّبة (سابقة، لاحقة) غير المتنافرة التي يمكن أن تجتمع في كلمة عربية سليمة واحدة هو بالتأكيد أقل من حاصل ضرب عدد السوابق في عدد اللواحق، وبالتالي فإن عدد هذه الأزواج ≈ (٣٠٠ × ٥٥٠) ÷ ع؛ حيث يمكن أن نعتبر «ع» معاملاً للتنافر، وتقدِّر بعض الدراسات الحاسوبية وكذلك بعض استقراءات السمات الأساسية لِلَّواصق أن قيمة «ع» تقع بين الثمانية والعشرة، وعلى ذلك نستطيع أن نقدِّر أن عدد هذه الأزواج الأزواج ≈ (٣٠٠ × ٥٥٠) ÷ ٩≈ ١٨,٠٠٠ زوجًا.

وبما أن كافة الجذوع الممكن توليدها تكاد تنقسم بالتساوي بين اسمية وفعلية، فإن العدد الكلي للكلمات الناتجة من التحام الجذوع بأزواج اللواصق يمكن تقديره تقريبيًّا

تمحيص وغربلة كل الروايات - لاستبعاد المكرَّر والمصَحَّف ... إلخ - تثمر عن عدد بين أربعة آلاف وخمسة آلاف جِذر مصرَّف.

١- ويغطي هذا العدد الحالات الأساسية إضافةً لكافة حالات الإعلال والإبدال والحذف.

٢- عدد المشتقات الممكنة من صب الجذر الواحد المصرَّف في كل الصيغ الصرفية (بكل حالات الإعلال والإبدال والحذف) يساوي فقط عدد الصيغ الصرفية الأساسية، ولذلك استخدمنا هذا العددَ في حاصل الضرب.

عبر حاصل الضرب (١٨,٠٠٠× ٥٠٠,٠٠٠) ÷٢= ٤,٥٠٠,٠٠٠,٠٠٠ كلمة عربية؛ أي «أربعة آلاف وخمسمئة مليون» كلمة، أو بتعبير آخر «أربعة ونصف بِلْيون» كلمة. ويشكل هذا العدد المهول حجم «فضاء المفردات» (Vocabulary Space) العربية الذي لا بُدَّ أن تنتمي إليه أية كلمة عربية سليمة.

وعلى الجانب الآخر لتقدير حجم حصيلة المفردات على السماع، فمن الواضح أن هذه الحصيلة السماعية تحتل مجرد حيز جزئي من فضاء المفردات العربية (الحصيلة القياسية)، ولكننا نحتاج من أجل تحديد هذه الحصيلة السماعية إلى تجميع ثم دراسة ما كُتِبَ وما قِيلَ بالعربية عبر الأبعاد الأربعة الأساسية التالية:

أ. البعد الزمني: حيث أنه كلما مر الزمن، جَدَّ استخدام كلمات من فضاء المفردات العربية لم تكن مسموعة من قبل[1]؛ ومن أمثلة هذه الكلمات العربية التي انتقلت إلى الحصيلة السماعية حديثًا؛ «اكتِئاب»، و«حاسوب»، و«ثَلَّاجَة»، و«الرَّسْمَلَة»، و«الإِشْتِرَاكِيَّة»، ... إلخ.

ب. البعد الجغرافي: حيث يشيع استخدام كلمات عربية سليمة في مناطق عربية معيَّنة بينما هي مهجورة أو نادرة الاستخدام في باقي المناطق، فعلى سبيل المثال تستخدَم في المغرب العربي دون غيره من المناطق العربية كلماتٌ عربيةٌ سليمةٌ مثل؛ «تَظاهُرة» (مناسبة عامَّة مثل المعرض أو المهرجان)، و«مُفَتِّشِيَّة» (هيئةُ أو سلطةُ التفتيش)، و«الفارط» (صفة تُسْتَخْدَم مع الأزمنة كالشهر والسنة بمعنى السابق أو الفائت)، و«دَرَكِيّ» (شُرَطِيّ)، «حَرَّاقة» (مهاجِرون غير شرعيين في بلاد أجنبية) ... إلخ.

ج. البعد الموضوعي: حيث تُسْتَخْدَم كلمات عربية سليمة كاصطلاحات في مجالات/ موضوعات بعينها بينما يقل أو ينعدم ذكرها فيما سواها من مجالات/ موضوعات؛ ومن أمثلة هذه الكلمات «الأَيْض» (في علوم الأحياء والكيمياء العضوية)، و «الائْتِكال» (في علوم الكيمياء وهندسة المعادن)، و«الاحْتِباك» (في مجال البلاغة)، و «مُجَحْفَل» (في العلوم

١- هذا بالطبع إضافةً إلى تلك الكلمات المسموعة بالفعل والتي تكتسب دلالات جديدة مع مرور الزمن.

العسكرية) ... إلخ.

د. البعد الأسلوبي: حيث تختلف الكلمات المستخدمة مع أسلوب الكلام، ففي الأسلوب السَّرْدِيّ يغلب مثلاً إسناد الأفعال للغائب؛ مثل «قالَ»، «قالوا»، «صَرَّحَ»، «أَكَّدَتْ» ... إلخ، كما يندر الإسناد للمتكلِّم والمخاطَب، ويندر استخدام صيغة الأمر، ويندر استخدام أدوات النداء، ... إلخ، أما في الأسلوب الحِواريّ يكثر إسناد الأفعال للمخاطَب، كما يمكن مصادفة الإسناد للمتكلم، ويمكن مصادفة أفعال الأمر، ... إلخ، وكذلك يتميز الأسلوب الخَطَابي بمفردات خاصة، ... وهكذا.

نحتاج إذن لجمع ما قيل وما كتب بالعربية على امتداد هذه الأبعاد الأربعة من أجل استخلاص حصيلة المفردات العربية السمعية حصرًا وعدًّا، وفي حين أن حصر وعد المفردات هو أمر متيسر حاسوبيًّا، فإن مهمة الجمع غير هينة أبدًا ليس فقط لضخامة الحجم بل أيضًا لأن الكثير من نطاقات الأبعاد السابق ذكرها غير مغطاة رقميًّا بشكل كافٍ؛ وذلك مثل اللغة التراثية، والأسلوب الحواري.

> يمكن طرح عمليات الجمع والحصر والعد السابق ذكرُها كمشروع كبير مقترح يتعاون فيه اللغويون والحاسوبيون بهدف حصر حصيلة المفردات اللغوية العربية السماعية وعَدِّها، حيث أن ذلك العمل لم يتم بشكل مستوفٍ حتى تاريخ كتابة هذه السطور.

وأيًّا ما كان التقدير الدقيق لحجم الحصيلة السماعية – التي تتراوح تقديراتها «الجُزَافيّة» بين الواحد في الألف والواحد في المئة من حجم فضاء المفردات العربية – فإنه يمكن بثقة تقرير عدد من الحقائق الهامة:

أ. إن فضاء المفردات العربية ذو حجم هائل.

ب. إن حصيلة المفردات العربية السماعية ما زالت تمثل جزءًا ضئيلاً من فضاء المفردات العربية.

ج. إن حصيلة المفردات العربية السماعية بالرغم من ذلك ذات حجم ضخم (يقدر بملايين الكلمات حسب أكثر التقديرات تواضعًا) يفوق حجم حصيلة المفردات السماعية في لغات العالم الحية الكبرى، ولذلك فإن تلك الحصيلة تحتاج عادةً إلى

مزيدٍ من العمل (مقارنةً بنظيراتها في لغات العالم الكبرى الأخرى) من أجل التعامل معها بفَعَالِيَةٍ في مختلف تطبيقات حوسبة اللغة العربية.

د. إن المساحات غير المستخدمة في فضاء المفردات العربية كبيرة جدًّا وتُشكِّل احتياطيًّا لفظيًّا إستراتيجيًّا يمكِّن اللغة العربية من الاستجابة الذاتية المرنة للاحتياجات الاصطلاحية الراهنة والمستقبلية في مختلف مجالات النشاط الإنساني، وذلك في تواصل مطرد بسلاسة مع التاريخ العريق الممتد اللغة.

٦- تعريف التحليل الصرفي الآلي وتطبيقاته

تتعلق دِلالات كلمة «مُورْفُولُوجِي» (Morphology) بمظهر وشكل الشيء/ الكائن وكذا قدرتِه على التشكُّل ودراسة تركيبه من الأجزاء المكونة له ... وما إلى ذلك، وتُسْتخدم الكلمة بشكل اصطلاحي في عدة مجالات مختلفة؛ ففي علم الأحياء على سبيل المثال يعني الاصطلاح دراسة مظهر وشكل وتركيب الكائنِ الحيِّ هو وأعضائِه[1]، أما في مجال علوم اللغة فإن الاصطلاح يعني دراسة بنية الكلمات وآلية توليدها من مكوِّناتها الأساسية وهو ما يُعْرَف في العربية بعلم الصَّرْف.

ومن خلال ما استعرَضَتْه الأقسامُ السابقة من هذا الفصل يتبين الانتظامُ العالي، وكذلك القدرةُ التوليدية الهائلة لنظام الصرف العربي عبر صب الجذور في الصِيَغ الصرفية لإنتاج الجذوع ثم التحام هذه الجذوع مع اللواصق لتوليد آلاف الملايين (٤,٥٠٠ مليون) من الكلمات وذلك من خلال استخدام عدة آلاف قليلة فقط (٦,٦٠٠) من المكوِّنات الصرفية، ويمكن تقديرُ مُعامِلٍ للتوليد الصرفي بخارج قسمة العدديـن ≈ ٦٨٢ مليون، وهو معامل هائل بكل معايير اللغات المعروفة!

وتُعَدُّ هذه القدرة التوليدية الصرفية العظيمة في اللغة العربية ميزةً كبرى في صالح تعلُّم اللغة العربية وكذلك ميزةً كبرى لتطبيقات معالجتها حاسوبيًّا كما سوف يلي توضيحه لاحقًا في هذا القِسْم، ولكن علينا قبل ذلك الانتباه إلى أن التحليل - وليس التوليد - هو المسألة الصرفية الطبيعية في معالجة المفردات حاسوبيًّا؛ ذلك أن الشكل

١- بينما يُعْنَى «الفِسْيُولُوجِي» Physiology في علم الأحياء بالوظائف الحيوية للكائن ولأعضائه.

اللغوي الذي ينتجه البشر بشكل طبيعي والمستهدَف معالجتُهُ حاسوبيًّا هو النص (أو الكلام) والذي هو عبارة عن سلسلة من الكلمات المولَّدة بالفعل، والمطلوب من هذه المعالجة هو تفكيك/ تحليل كلٍّ من هذه الكلمات إلى مكوِّناتها البنائية الأساسية.

ونستطيع ببساطة وبناءً على ما خُتِمَ به القِسْمُ قبلَ السابق من هذا الفصل أن نضع التعريف الوظيفي لأي «محلِّل صرفيّ عربيّ» كنظامٍ كما في الشكل رقم (٣-٤) أدناه:

كلمة عربية ← **تحليل صرفي** ← (نوع الكلمة: السابقة، الجذر، الصيغة الصرفية، اللاحقة)

الشَّكل ٣-٤: التعريف الوظيفي لنظام التحليل الصرفي العربي.

ونكرر هنا إلى أنه في العادة يكتفي المحلل الصرفي بإخراج «أكواد» كلٍّ من المكوِّنات الصرفية الأساسية الخمس التي يمكن عَبْرَها الوصول إلى التوصيف الحاسوبي لكلٍّ من هذه المكونات في برمجيات التحليل الصرفي العربي. ويعرض الجدول رقم (٣-١) أدناه عينةً من مخرجَات أحد برمجيات التحليل الصرفي العربي[١]:

عند هذه النقطة وقد اتضحت البنية الصرفية للمفردات العربية وصار لدينا تعريف محدد للتحليل الصرفي العربي، من الطبيعي أن نتساءل عن أهمية حوسبة التحليل الصرفي العربي وفوائده التطبيقية.

وفي الواقع فإن انتظام عملية التوليد الصرفي للكلمات مع الارتفاع الهائل في مُعامِله هما ميزتان كُبْرَيَان لِلُّغة العربية تضفيان مزيدًا من الأهمية على دَوْر الصرف في كلٍّ من استخدامها البشري وكذلك معالجتها الحاسوبية، حيث تشكلان معًا ركيزةً لغوية للناطقين بالعربية ومتعلميها، كما تمثلان محورًا للعديد من تطبيقات المعالجة الحاسوبية للغة العربية.

الكلمة	نوعها	السابقة	الجذر	الصيغة الصرفية	اللاحقة
فَمَا	جامِدة	فَـ ٢	اَلَّذِي ٨٧	مَا ٤٨	– ٠

١- وهو الموصوف في المرجعين رقم [١٥] ورقم [١٦] فى قائمة المراجع نهاية هذا الكتاب.

الكلمة	نوعها	السابقة	الجذر	الصيغة الصرفية	اللاحقة
تَتَنَاوَله	مصرَّفة منتظِمة	تـ ٨٦	ن و ل ٤٠٧٧	تَفَاعَلَ ١٧٦	ـه ٨
اَلْكِتَابَات	مصرَّفة منتظِمة	الـ ٩	ك ت ب ٣٣٥٤	فِعَال ٦٨٤	ـات ٢٧
اَلْعِلْمِيَّة	مصرَّفة منتظِمة	الـ ٩	ع ل م ٢٧٥٤	فِعْل ٨٤٢	ـيَّة ٢٨
مِنْ	جامِدة	– ٠	مِنْ ٦٣	مِنْ ١١٨	– ٠
مَوَاضِيع	مصرَّفة منتظِمة	– ٠	و ض ع ٤٣٣٩	مَفَاعِيل ٩٣	– ٠
مُتَّخَذَة	استثناء صرفيّ	– ٠	أ خ ذ ٣٩	مُتَّخَذ ١٣	ـة ٢٦

الجدول ٣-١ : أمثلة على التحليل الصرفي لبعض الكلمات العربية.

فالمَلَكات اللغوية للناطقين باللسان العربي تتشكل وفي أُسُسِها أن السيطرة الذهنية على الحصيلة اللغوية ذات الحجم الضخم لا تقوم فقط على حفظ واستظهار مفرداتها بمعزل عن بعضها البعض بل تقوم إلى حد بعيد على الترابط البنائي والدلالي المتين بين هذه المفردات؛ فمن الواضح على سبيل المثال أن كلمات مثل{«عِلْم»، «عُلُوم»، «تَعَلم»، «تَعَلُّم»، «مُتَعَلِّم»، «مُعَلِّم»، «عَالِم»،...}- والتي تشترك في الجِذر «ع ل م» – تدور جميعها حول معنًى مشترك عام (وهو المعرفة والفهم) وذلك على الرغم من أن لكلٍّ منها دِلالة تفصيلية مختلفة[1]، كما أن كلمات مثل {«اسْتِعْلام»، «اسْتِعْطَاف»، «اسْتِحْضَار»، «اسْتِنْصَار»، «اسْتِكْثَار»، «اسْتِطْلاع»،...}- والتي تشترك في القالب

١- انتبه علماء اللغة العرب منذ وقت باكر (راجع مثلاً «ابن فارس» في قاموسه «مقاييس اللغة» وهو المرجع رقم [١] في قائمة المراجع نهاية هذا الكتاب) إلى أن الجذوع المشتقَّة من كل جذر مصرَّف تدور حول عدد قليل من المعاني الكلية (واحد أو اثنين أو ثلاثة، وفي حالات نادرة يزيد عدد هذه المعاني الكلية عن ذلك). وقد جرى استثمار هذه الحقيقة الهامة في الأعمال الحديثة لحوسبة «الدِلالات المعجمية العربية» وهو ما سيُفْرَد له الفصل الأول من الباب السادس «التحليل الدلالي» في هذا الكتاب.

ولتكملة الإفادة من المثال المضروب على جذر «ع ل م»، فإن له – حسب «مقاييس ابن فارس» أيضًا – معنى كليًا آخَر وهو (الأثَر يتميز عن سواه) ومن مشتقات الجذر التي تدور حول هذا المعنى الكلي {«عَلَم»، «أَعْلام»، «مَعْلَم»، «مَعَالِم»، ...}.

الصرفي «اسْتِفْعَال» – تحمل جميعها ملمحًا دِلاليًّا عامًّا (وهو طلب الشيء والسعي لتحصيله)، كما أن اللواصق تحمل دِلالات ووظائف نحوية محدَّدة. إذن، فإن تعلُّم عدد قليل نسبيًّا من اللَّبِنات الصرفية (عدة آلاف) مع إتقان ضوابط التفاعلات بين هذه اللبنات يمكِّن المرء من السيطرة الدلالية والبنائية على الحصيلة اللغوية العربية هائلة الحجم (بين الملايين والبلايين)[١] مما ييسر عليه القراءة والفهم وكذلك التعبير والكتابة.

أما بالنسبة لمعالجة اللغة العربية حاسوبيًّا فإن أهمية الحصول على محلل صرفي عربي آلي تكمن في اختصار العبء الثقيل للتعامل مع العدد الهائل من حصيلة مفردات اللغة العربية عبر تفكيكها إلى عدد محدود من اللَّبِنات الصرفية العربية الأولية. ومن حيث الاعتماد على التحليل الصرفي العربي الآلي؛ فإن بعض التطبيقات للمعالجة الحاسوبية لِلُّغة العربية محورُها الأساسي هو التحليل الصرفي العربي، بينما بعض التطبيقات الأخرى يكون التحليل الصرفي العربي الآلي وسيلةً ضروريةً أو مساعِدةً لإنجازها.

ومن بين التطبيقات الحاسوبية التي تنتمي للنوع الأول؛ نذكر على سبيل المثال لا الحصر:

أ. دعم وظائف البحث الاشتقاقي العربي في محركات البحث النصي: فالمستخدم عندما يبحث عن كلمةٍ ما في نطاقٍ نصيٍّ معين؛ ولتكن كلمة «اعتماد» على سبيل المثال، فإنه كثيرًا ما يرغب في الحصول ضمن نتائج بحثه على مواقع كلماتٍ مثل «اعتمادات»، «فالاعتماد»، «واعتمادهم»، «الاعتمادية» ... وكل الكلمات الأخرى المشتركة في نفس الجذع والواردة في نطاق البحث (ويسمى ذلك «نمط البحث على مستوى الجذع»)، وقد يود المستخدم توسيع نتائج البحث حتى يحصل إضافةً على ما سبق على مواقع كلماتٍ مثل «يعتمدونه»، «معتمد»، «عماد»، «العمود» ... وكل الكلمات الأخرى المشتركة في نفس الجذر والواردة كذلك في نطاق البحث (ويسمى ذلك «نمط البحث على مستوى «الجذر»)، ... وهكذا.

١- كما جرى تفصيلُه في القسم السابق من هذا الفصل؛ فإن تقدير حجم الحصيلة اللغوية العربية القياسية قد يصل إلى أربعة ونصف بليون كلمة، بينما تتراوح التقديرات «الجزافية» لحجم الحصيلة اللغوية العربية السماعية بين بضعة ملايين (كحد أدنى) وعشرات الملايين (كحد أقصى) من الكلمات.

ب. الربط المعجمي العربي الآلي: فالقارئ للنص العربي إلكترونيًّا ربما يحتاج إلى شروح لبعض المفردات التي تصادفه، ولكن مداخل المعاجم العربية هي مواد الجذور (وليست الكلمات – نظرًا لاتساع الحصيلة اللغوية العربية السابق ذِكرُه) والتي يتطلب الوصولُ إليها استخدامَ التحليل الصرفي الآلي، ويتيح هذا الاستخدام أيضًا فرصة إضافة مداخل أخرى لشرح الصيغ الصرفية واللواصق مما يستوفي شرح المفردة من كافة زواياها.

ج. التحليل الدِّلالي المعجمي العربي: يكاد يكون من المستحيل تغطية كل المفردات العربية ككلماتٍ تامةٍ – وذلك مرةً أخرى بسبب ضخامة حجم حصيلتها – بشكلٍ كافٍ في قاعدة بيانات موسَّعة ترتبط فيها كل مفردة بما يناسبها من حقولٍ دِلاليةٍ (معانٍ عامَّة) كما ترتبط عبر علاقات دِلالية ببعضها البعض، ولذلك تُستبدَل الكلمات التامة بتركيباتٍ صرفية أكثرَ مرونة منها تسمح بتغطية مفردات اللغة العربية بنسبة عالية في أية قاعدة بيانات للربط الدِّلالي، كما يُستخدَم التحليل الصرفي الآلي للعبور من كلمات النصوص المطلوب دراستها دِلاليًّا إلى هذه التركيبات الصرفية(١).

د. البرمجيات التي تساعد في تعلُّم اللغة العربية: وخصوصًا تلك المعنية بالهجاء/ الإملاء، وعِلْم الصرف.

وكذلك العديد من التطبيقات الأخرى التي يشكِّل التحليل الصرفي العربي الآلي محورَها الأساسي.

أما على جانب تطبيقات المعالجة الحاسوبية لِلُّغة العربية التي يمثِّل التحليل الصرفي العربي الآلي وسيلةً ضروريةً أو مساعِدةً لإنجازها فمن أمثلتها؛ «تشكيل النَّصّ العربي» (Arabic Diacritization)[٢] (حيث تحتوي نواتج التحليل الصرفي لكل كلمة على «سابقة» + «صيغة صرفية» + «لاحقة» معًا، وبعد تطبيق التأثيرات البنيوية لالتحامها على التشكيل الصوتي للكلمة ما عدا الحركة الصوتية الإعرابية أحيانًا)[٣]،

١– للمزيد حول ذلك الموضوع يمكن الرجوع إلى الفصل الخامس «التحليل الدلالي» في هذا الكتاب.

٢– ويُقْصَد بذلك النظم الآلية لاستنباط علامات الضبط الصوتي لكلمات نصٍّ عربيٍّ ما.

٣– وهو بدوره يدخل في نظم معالجة لغوية أخرى؛ من بينها على سبيل المثال «تخليق الكلام المنطوق من النص العربي

وكذلك التدقيق الإملائي في النص العربي (لاكتشاف مواضع الأخطاء الإملائية و/ أو تصحيحها)، ...، إلخ.

ويجادل البعض بعدم الحاجة إلى عملية التحليل الصرفي في التطبيقات التي يمثِّل التحليل الصرفي العربي الآلي وسيلةً لإنجازها، وبأن مثل هذه التطبيقات يكفيها الاقتصار على التعامل مع حصيلة المفردات العربية التامة بشكل مباشر عبر استقرائها من مدوَّنات نصية عربية موسَّعة وبتوظيف أساليب حاسوبية إحصائية وبالتعويل على القدرات الكبيرة (والمتصاعدة) للأجيال الحالية من الحاسبات الرقمية، ولكن أرفع الأبحاث المدققة بينت حديثًا أن تغطية حصيلة المفردات العربية لا يمكن تحقيقها حاسوبيًّا بشكل مُرْضٍ وأن أكفأ وأكمل الحلول العملية المتاحة تتكون من مزيج/ هجين من كلا التوجهين[1].

٧- التركيب الصرفي معبرًا لحوسبة التحليل الصرفي

كيف يجري تنفيذ التحليل الصرفي العربي حاسوبيًّا؟

من أفضل المقارَبات الرياضية المحْكَمة في علم الذكاء الاصطناعي التي يمكن استعارتها لإجراء التحليل الصرفي حاسوبيًّا بكفاءةٍ أسلوبٌ يُعْرَف بِاسْم «التقييد المتعاقِب» (Successive Constraining)، وتقوم فكرته الأساسية على آليتين تعملان بالتعاقُب؛ أولاهما تطرح فرضيات عن كل الحلول الممكِنة لأجزاء متتالية (أو مستويات متصاعدة) من المسألة قيد البحث ومن ثَمَّ تضيفها إلى فضاء الحلول الممكنة، وثانيتهما تختبر هذه الفرضيات مقابل القيود التي لا بُدَّ أن يلتزم بها أي حل مقبول للمسألة ومن ثَمَّ تحذفها من فضاء الحلول الممكنة، وتعمل هاتان الآليتان بالتعاقب إلى أن تُسْتَنْفَدَ

المكتوب»؛ وهذا الأخير بدوره يعتبر أساسًا تقوم عليه عدة تطبيقات مثل «تمكين المكفوفين من استخدام الكمبيوتر عبر قراءة محتويات شاشته»، وأنظمة «الاستجابة التفاعلية الصوتية IVR» ... إلخ.

١- ومن أمثلة تلك الدراسات الحديثة ما جاء في الورقة البحثية الآتية:

Rashwan ,M ,.Al-Badrashiny ,M ,.Attia ,M ,.Abdou ,S ,.Rafea ,A ,.A Stochastic Arabic Diacritizer Based on a Hybrid of Factorized and Un-factorized Textual Features ,IEEE Transactions on Audio ,Speech ,and Language Processing) TASLP (http//:www.SignalProcessingSociety.org/Publications/Periodicals/TASLP ,Vol – 19 .Issue ;1 pp 166-175 .http//:ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber ,5428861=Jan.2011 .

أجزاء/ مستويات المسألة فتكون فئة الحلول الممكِنة للمسألة بكاملها عندئذٍ هي تلك التي نجت من الحذف وبقيت في فضاء الحلول.

ولإسقاط «التقييد المتعاقِب» على مسألة التحليل الصرفي العربي[1]، فإن علينا اتباعَ الخطوات التالية:

أ. تضاف أية سابقة ينطبق هجاؤها على صَدْرٍ من هجاء الكلمة محلَّ الدراسة إلى فضاء الحلول. ((آلية طرح فرضيات)).

ب. تضاف أية لاحقة ينطبق هجاؤها على عَجُزٍ(٢) من هجاء الكلمة محلَّ الدراسة إلى فضاء الحلول. ((آلية طرح فرضيات)).

ج. من بين القائمتين اللتين حصلنا عليهما في الخطوتين السابقتين يتم توليد كل الأزواج المرتَّبة الممكنة (سابقة، لاحقة) ومن ثَمَّ يتم دفعُها إلى فضاء الحلول. ((آلية طرح فرضيات)).

د. يتم استبعادُ الأزواج المتنافرة من فضاء الحلول. ((آلية تقييد)).

هـ. لكل زوج مرتَّب متبقٍّ في فضاء الحلول من الخطوة السابقة؛ يتم استخلاص جذع الكلمة (الكلمة بدون السابقة واللاحقة) مع التعويض بالطبع عن التأثيرات البنيوية من التحام السابقة بالجذع وكذلك من التحام اللاحقة بالجذع. وتضافُ الجذوعُ الناتجةُ عن هذه العملية إلى فضاء الحل. ((آلية طرح فرضيات)).

و. تُستبعَد الجذوع المكرَّرة من فضاء الحلول. ((آلية تقييد)).

ز. لكلٍّ من الجذوع المتبقية في فضاء الحلول، يجري البحث في التوصيفات الحاسوبية لكلِّ الجذور والصيغ الصرفية في القاعدة اللغوية لبرنامج التحليل الصرفي عن كل الأزواج المرتَّبة على هيئة (نوع الكلمة: جذر، صيغة صرفية) التي يكون الهجاء الناتج من صب جذرها في صيغتها الصرفية مطابقًا لهجاء الجذع، وتُدْفَعُ فرضيات التحليل الصرفي للكلمة العربية محل الدراسة إلى فضاء

١- لمزيد من التفاصيل التقنية الدقيقة بهذا الخصوص يمكن الاطلاع على الفصل السادس من المرجع رقم [١٥] في قائمة المراجع نهاية هذا الكتاب.

٢- عَجُز الشيء هو مؤخرته أو أسفله أو خاتمته، والجمع أَعْجاز.

الحلول في صورتها القياسية (نوع الكلمة: سابقة، جذر، صيغة صرفية، لاحقة). «آلية طرح فرضيات».

ح. وبالطبع تُستبعَد من فضاء الحلول تلك الجذوع التي لا يمكن تكوينها بصب أي جذر في أية صيغة صرفية. «آلية تقييد».

ط. لكلٍّ من التحليلات الصرفية الموجودة في فضاء الحلول بعد الخطوة السابقة؛ يجري اختبار عدم التنافر بين السابقة والصيغة الصرفية وكذلك بين الصيغة الصرفية وبين اللاحقة، وتُستبعَد تبعًا لذلك التحليلات التي تحتوي على أي تنافر. «آلية تقييد».

ي. يتم تركيب كلمةٍ من كل تحليل متبقٍّ في فضاء الحل بعد الخطوة السابقة، ثم تُطابَق الكلمة الناتجة من التركيب مقابل الكلمة محل الدراسة من حيث الهجاء وكذلك من حيث الضبط الصوتي (إن وُجِدَ أيُّ ضبط صوتي في الكلمة المدخَلة محل الدراسة)، وتُستبعَد التحليلات الصرفية غير المطابقة من فضاء الحلول. «آلية تقييد».

ك. التحليلات المتبقية (الناجية) الآن في فضاء الحلول هي فئة كل التحليلات الصرفية الممكِنة للكلمة العربية محل الدراسة. «آلية تقييد».

وجدير بالملاحظة في كل هذه الخطوات أن الحصول على التحليلات الصرفية للكلمة يتم تدريجيًّا عبر التركيب الصرفي (وفق ما جرى عرضه في الأقسام السابقة من هذا الفصل) لفرضيات تحليلاتها الصرفية الجزئية، ثم اختبار تحقيق هذه الفرضيات للشروط الصرفية العربية واستبعاد ما لا يستوفي هذه الشروط، وبتكرار هذه العملية إلى أن تُغطِّي جسم الكلمة محل الدراسة يُشَكِّل ما يتبقى من هذه الفرضيات فئةَ الحلول الممكِنة لهذه الكلمة.

٨- الالتباس الصرفي وأساليب إزالته

حَسْبَ ما اخْتُتِمَ به القسمُ السابقُ فإن عملية التحليل الصرفي العربي حاسوبيًّا تُفْضِي إما إلى عدة تحليلات صرفية قياسية مُمْكِنة أو إلى تحليل صرفي وحيد أو ربما إلى عدم وجود تحليل صرفي عربي على الإطلاق، وهذا كله يعني على وجه العموم أن تطبيق

هذه العملية على أية كلمة عربية يُنْتِج تحليلاتٍ متعددةً وهو ما يُعْرَفُ اصطلاحًا باسم «الالتباس الصرفي» (Morphological Ambiguity).

ويعرض الجدول رقم (٢-٣) أدنـاه مثالاً مفصَّلاً على هذا الالتبـاس عنـد إجراء عملية التحليل الصرفي العربي حاسوبيًّا على الكلمة العربية الخام «بطيـن»؛ حيث تم تظليل كل تحليل ممكن من التحليلات الصرفية القياسية المتعددة لهـذه الكلـمـة بظـل رمـادي ذي كثافةٍ متناسبةٍ مع درجة الشيـوع السمـاعية لهذا التحليل[1].

وفي غياب آليةٍ تختار واحدًا فقط من بين التحليلات الصرفية المتعددة - وهو ما يُعْرَف اصطلاحًا بِاسْـم «فَكّ الالتبـاس الصـرفي» (Morphological Disambiguation) - فلا يمكن أن يُعتبَر هذا التحليل الصرفي العربي آليًّا ويظل ذا قيمه تطبيقية محدودة.

ويجدر بنا قبلَ مناقشة كيفية فك الالتباس الصرفي العربي أن نلقي نظرةً على الأسباب الرئيسية لهذا الالتباس والتي يمكن تلخيصها فيما يلي:

أ. افتقاد علامات التشكيل/ الضبط الصوتي - إلا فيما ندر - في الكتابة المعاصرة للنص العربي وذلك ارتكانًا على قدرة القارئ العربي المتمرس على استنباط هذه العلامات بنفسه مستفيدًا من عدة قرائن سوف نذكرها فيما يلي من نقاط.

بل إنه من الجدير بالذكر أن استنتاج علامات التشكيل الصوتي حاسوبيًّا لكلمات النص العربي الخام هو بحد ذاته هدف للتحليل الحاسوبي المعمَّق لِلُّغة العربية يُسْهِم فيه التحليل الصرفي العربي الآلي كما أشرنا إلى ذلك في القسم قبل السابق من هذا الفصل. وعلى الرغم من أن إضافة التشكيل الصوتي ولو جزئيًّا لبعض الكلمات قد يفك/ يقلل من الالتباس فيها؛ وذلك مثل «مستعمِر»، «مقدَّم»، «يقسِّم» ... ، فإن علينا الانتباه في الحالة العامة إلى عدم كفاية إضافة التشكيل الصوتي ولو كان كاملاً لإزالة الالتباس؛ وذلك مثل «فَرَضُوا»، «قَائِل»، «يَسِير»، «شَذَا» ... فقد يكون المثال الأول فعلاً ماضيًا من مادة الجذر (ف، ر، ض) أو

١- المحلل الصرفي العربي المستخدَم للحصول على التحليلات الواردة في الجدول المذكور هو ذلك الموصوف في المرجع رقم [١٥] (وخصوصًا الفصول من خمسة إلى سبعة) من قائمة المراجع نهاية هذا الكتاب.

فعلاً ماضيًا معطوفًا بالفاء من مادة الجذر (ر، ض، ي)، كما قد يكون المثال الثاني اسم فاعـل من مادة الجذر (ق، و، ل) أو من مادة الجذر (ق، ي، ل)، أما المثال الثالث فقد يكون فعلاً مضارعًا من مادة الجذر (س، ي، ر) أو صيغة مبالغة من مادة الجذر (ي، س، ر) أما المثال الرابع فقد يكون فعلاً ماضيًا أو اسمًا وكلاهما من مادة الجذر (ش، ذ، و).

ب. بينما تشتغل أية آلية للتحليل الصرفي على كل كلمة في النص بشكل منفرد، فإن الذهن البشري يستفيد في حكمه على أية كلمة من سياق النص حولها؛ فعلى سبيل المثال نستطيع كقارئين للعربية بكفاءة فك التباس كلمة «بطين» عندما ترد في السياقات الثلاث المختلفة الآتية (... يحمل الدم الموجود في البطين الأيمن من القلب ...)، و(... الذي يختلط بطين الأرض فيصير ...)، و(... أنه رجل بطين جسيم...).

ج. وفي قدراته تلك على فك الالتباس الصرفي يعتمد الذهن البشري على المعالجة المتزامنة لسياق الكلمة على مستويات عدة من التحليل اللغوي - وخصوصًا الدلالة والنحو - إضافةً إلى الصرف، ولسوء الحظ لا تزال أدواتنا الحاسوبية والرياضية قاصرة عن محاكاة هذه المعالجة اللغوية المتزامنة بكفاءة لحظية. وكمثال على ذلك نحلل السياق التالي (... فالكهرباء وقود يسير هذا النوع من السيارات.) حيث يستبعد القارئ العربي تفسير ما تحته خط على أنه الصفة «يَسِير» من مادة الجذر (ي، س، ر) في صيغة المبالغة «فَعِيل» وذلك لأن ما يتيحه السياق النحوي من موصوفات محتملة هو «فالكهرباء» أو «وقود» وكلاهما بعيد دلاليًّا عن وَسْمِه بصفة «اليُسْر»، كما يستبعد القارئ العربي أيضًا تفسيرَه على أنه الفعل اللازم «يَسِير» من مادة جذر (س، ي، ر) وذلك لأن اعتبار «هذا» كفاعل صريح يؤدي إلى اختلال دِلالة العبارة وبنائها حيث تفتقر الجملة الفعلية التي تقع في محل صِفَة للاسم قبلها إلى متعلق يربطها بموصوفها - كأن يقال (... وقود يَسِيرُ به هذا النوع ...) - كما أن اعتبار الفاعل ضميرًا مستترًا يعود على «فالكهرباء» أو «الوقود» غير مناسب دلاليًّا حيث إن أيًّا منهما لا «يَسِير»، ويُستبعَد كذلك التفسير «يُسَيَّر» لأسباب مماثلة، ...، وربما لا يتبقى

لدى القارئ في نهاية المطاف احتمال قوي يحقق صحة الكلام صرفيًّا ونحويًّا ودلاليًّا في آن واحد سوى «يُسَيِّر»[1].

د. إضافةً إلى سياق النص الذي ترد فيه الكلمة محل الدراسة والتحليل، يوجد سياق آخر لا يقل أهمية عنه في فك التباسها؛ ألا وهو «السياق خارجَ النص» ونقصد به حقائق العالم وتفاعلاته وملابسات الأحداث المعاصرة للكلام ... إلخ، ومرة أخرى ومع الأسف فإنه لم يتم إلى الآن بناء قاعدة معرفية حاسوبية تصف بصورة مكتملة العالم المشترك بين عموم البشر[2] ناهيك عن القدرة على حوسبة كل سياق خارجي خاص بفئة معينة من البشر كالمحامين، أو مهندسي الاتصالات، أو الأدباء، أو عمال البناء، أو سكان حي معين في بلدةٍ ما، ... إلخ.

وقد تبينت أهمية هذا السياق الخارجي أثناء مناقشة المثال في النقطة السابقة، ولمزيد من التبيان نحلل السياق التالي (... وقد تزينت البليدة لهذا الحدث الهام ...) فإذا كان القارئ يعرف أن السياق الخارجي للنص يشير إلى بلدنا العزيز «الجزائر» (كأن يرد مثلاً في جريدة تصدر في الجزائر، ... إلخ) وكان يعرف أيضًا أن «الْبُلَيْدَة» هي إحدى مدنها التي تتأهب لاستقبال مناسبة هامة فإنه يستطيع تقرير أن ما تحته خط ليس كلمة «الْبَلِيدَة» (التي قد تصف بنتًا أو امرأةً تتزين زينةً حقيقة) وإنما هي كلمة «الْبُلَيْدَة» (بافتراض علم القارئ أن فعل التزيُّن يمكن أن يُسْنَد إلى اسم علم على مدينة كاستخدام مجازي بلاغي عبر الاستعارة المكْنِيَّة).

الكلمة	نوعها	السابقة	الجذر	الصيغة الصرفية	اللاحقة
بَطِين	مصرَّفة منتظِمة	– ٠	ب ط ن ٣٥٢	فَعِيل ٦٧٣	– ٠
بُطَيْن	مصرَّفة منتظِمة	– ٠	ب ط ن ٣٥٢	فُعَيْل ٧٨٩	– ٠

١- لمزيد من التفاصيل حول حوسبة المعالجة الدلالية يرجى الرجوع إلى الباب السادس من هذا الكتاب حول «التحليل الدلالي».

٢- تسمى مثل هذه القاعدة المعرفية «أنطولوجيا Ontology» وفي الفصل الخامس «التحليل الدلالي» في هذا الكتاب نستعرض هذا المفهوم ونناقش أهم مكوناته وكذلك أهميته لحوسبة التحليل الدلالي.

الكلمة	نوعها	السابقة	الجذر	الصيغة الصرفية	اللاحقة
بَطَّيْن	مصرَّفة منتظِمة	- ٠	ب ط ط ٣٤٨	فَلّ ٨٢٠	ـيْنِ ٨٠
بـِطِين	مصرَّفة منتظِمة	بـ ١٥	ط ي ن ٢٥٦٣	فِعل ٨٤٧	- ٠
بـِطِّين	مصرَّفة منتظِمة	- ٠	ب ط ن ٣٥٢	فِعِّيل ٦٧٣	- ٠
بُطَيِّن	مصرَّفة منتظِمة	- ٠	ب ط ن ٣٥٢	فُعَيِّل ٧٨٨	- ٠
بـِطَيْن	مصرَّفة منتظِمة	بـ ١٥	ط ي ن ٢٥٦٣	فَعْل ٨١٩	- ٠
بُطَّيْن	مصرَّفة منتظِمة	- ٠	ب ط ط ٣٤٨	فُلّ ٨٣٤	ـيْنِ ٨٠
بـِطَّيْن	مصرَّفة منتظِمة	- ٠	ب ط ط ٣٤٨	فِلّ ٨٤٣	ـيْنِ ٨٠
بـِطِيَن	مصرَّفة منتظِمة	بـ ١٥	ط ي ن ٢٥٦٣	فِعَل ٨٥٠	- ٠
بـِطُيَّن	مصرَّفة منتظِمة	بـ ١٥	ط ي ن ٢٥٦٣	فُعَّل ٧٣٩	- ٠
بـِطَيِّن	مصرَّفة منتظِمة	بـ ١٥	ط ي ن ٢٥٦٣	فَعِّل ٦٧٥	- ٠

الجدول٣-٢: مثال على الالتباس الصرفي عند التحليل الصرفي الآلي لكلمة «بطين».

ومع بروز إشكالية الالتباس على المستوى الصرفي (وعلى باقي مستويات اللغة) كأمر لا فِكاك منه بكل تحدياته التي عرضناها عالِيَه، فقد توجه المشتغلون بالمعالجة الحاسوبية لِلُّغات الحية إلى استعارة وتطويع أساليب التعلم الحاسوبي التي أثبتت فَعَالِيَّتَها في التعامل مع مسائل صعبة وحيوية في صناعة المعلومات والاتصالات الحديثة مثل «معالجة الإشارات» (Signal Processing)، و «التعرف على الأنماط» (Pattern Recognition) ... إلخ. وبصفة عامة فإن أساليب «التعلم الحاسوبي» ((Machine

Learning تطبَّق كأنسب ما يكون لمقارَبة المسائل التي لا تُعْرَف لها أو حيث يتعذر الحصول لها على «حلول بقواعد مُحْكَمة»(Closed-Form Solutions) ، وتنطلق هذه الأساليب - على اختلافها - من مبدأ إمكانية التعلم عبر تكرار التعرض الغزير للأمثلة الصحيحة والأمثلة الخاطئة أو عبر تكرار التعرض الغزير للأسئلة وأجوبتها حول جزئيات المسألة المطلوب مقاربتها؛ فالطفل مثلاً قد يجيد القراءة السليمة للغة العربية قبل أن يتعلم أسس وقواعد اللغة من صرف ونحو ودلالة ... إلخ وذلك عبر تقليد الأكبر سنًّا وعبر محاولات الصواب والخطأ مع تصويب الأخطاء حيث يتحسن الأداء مع معاودَة ذلك كله مرارًا وتكرارًا.

ويرتكز بصفة عامة تفعيل تلك الأساليب رياضيًّا وحاسوبيًّا لهذا المبدأ على استقراء السياق الاحتمالي للكلمة عِوَضًا عن سياقها اللغوي (النحوي-الدلالي) بُغْيَةَ الوصول إلى حساب الاحتمال الرياضي لوقوع كل تحليل ممكن للكلمة بين ما يسبقه وما يلحق به من كلمات ومن ثَمَّ ترجيح التحليل صاحب أكبر احتمال رياضي، وتستلزم عملية إجراء الحساب هذه تكوين نموذج احتمال رياضي يحاكي كل تتابُعات الوحدات اللغوية كما تحدث في الواقع الحقيقي لاستخدام اللغة. وفي حين أن المرجع رقم [٢١] من قائمة المراجع نهاية هذا الكتاب يدرس باستفاضة أساليب التعلم الحاسوبي المختلفة بتفاصيلها الرياضية المعمَّقة، فإن المرجع رقم [٢٠] من نفس القائمة يستفيض في تطبيق بعض هذه الأساليب على معالجة اللغات الحية.

وأيًّا ما كانت الأساليب الرياضية والحاسوبية المختارة لبناء مثل هذا النموذج الاحتمالي، فلا بُدَّ لها من مادة لغوية تملأ وعاءً واسعًا من الأسئلة (الكلمات الخام على سبيل المثال) مع أجوبتها (التحليل الصرفي الصحيح للكلمة على سبيل المثال) مثلما يقتضي التعلم الحاسوبي الموجَّه، ويُطْلَق على عملية تشغيل هذه الأساليب الرياضية على المادة اللغوية لبناء النموذج الاحتمالي اسم «التدريب» (Training) كما يطلق على الوعاء الواسع الممتلئ بالمادة اللغوية اسم «المَوْرِد اللغوي» (Language Resource)، وبطبيعة الحال فإن النموذج الاحتمالي سوف يحمل الخصائص الإحصائية للمورد اللغوي الذي تَكَوَّنَ منه، والذي عليه بدوره أن يحمل خصائص الظاهرة اللغوية التي يعبر عنها وهو ما يُسْلمنا إلى اعتبارات وشروط بناء الموارِد اللغوية في القسم التالي.

٩- بناء الموارِد اللغوية اللازمة لفك الالتباس الصرفي

من أجل نجاح عملية فك الالتباس اللغوي (سواء كان صرفيًّا أو غير ذلك) فإن النموذج الاحتمالي الرياضي للظاهرة اللغوية محل الدرس لا بُدَّ أن يحمل قدرَ الإمكان ذاتَ الخصائص الإحصائية[١] لهذه الظاهرة اللغوية، ومن أجل تحقُّق ذلك لا بُدَّ أن يحمل المورد اللغوي بدوره قدرَ الإمكان ذاتَ الخصائص الإحصائية للظاهرة اللغوية، ومن أجل ذلك يتوجب أن يتوافر فيه الشرطان الآتيان:

أ. أن تُختار المادة الخام (سواء أكانت نصوصًا أم خلاف ذلك) في المورد اللغوي بحيث تُشكِّل عينةً حسنة التمثيل للظاهرة اللغوية، ويعني ذلك أن تغطي هذه العينة الأبعاد الأربعة لفضاء التنوع اللغوي السالف ذكرها في القسم الخامس من هذا الفصل (وهي البعد الزمني، والبعد الجغرافي، والبعد الموضوعي، والبعد الأسلوبي) بحيث تتناسب أوزان تمثيل كل حيز في هذا الفضاء في تلك المادة الخام قَدْرَ الإمكان مع أهميتها في التطبيق الذي سيُوظَّف فيه النموذج الاحتمالي الناتج من التدريب على هذا المورد اللغوي.

ب. أن تُجْرَى عملية عَنْوَنة المادة اللغوية الخام لإنتاج المورد اللغوي بدقة عالية، ويُقْصَد بعملية «العنونة» (Annotation)[٢] إلحاقُ التحليلات اللغوية المناسبة (كالتحليل الصرفي على سبيل المثال) بكل وحدة لغوية في المادة الخام لتُكوِّن جميعُها المورد اللغوي. وتنبع ضرورة الدقة العالية في هذا المقام من أن المورد اللغوي الناتج هو مادة التدريب (أيًّا كانت أساليب التدريب) لإنتاج النموذج الاحتمالي الرياضي الذي تعتمد عليه آليات فك الالتباس، فإن كان المورد رديء الجودة فالنموذج الاحتمالي الناتج عنه هو بالتأكيد معيبٌ

١- المقصود بالخصائص الإحصائية للظاهرة اللغوية هو التكرارات النسبية لوحداتها اللغوية في واقع تطبيق هذه الظاهرة اللغوية.

٢- تشيعُ أيضًا لفظة «توسيم» لترجمة المصطلح «Annotation».

وبالتبعية يصير أداءُ آليةِ فك الالتباس المعتمِدةِ عليه رديئًا، ويُعَدُّ هذا تجليًا لمبدأ في الرياضيات الحاسوبية مؤداه أن «معالجة النفايات المدْخَلة لا تُخْرِجُ إلا النفايـات» (Garbage In – Garbage Out) ويعرف بالاختصار «GIGO».

ومن النماذج الجيدة على بناء الموارِد اللغوية العربية الحديثة ما جاء في الورقة المشار إليها برقم [٢٢] فى قائمة المراجع نهاية هذا الكتاب والتي تناولت عدة أصناف من الموارِد اللغوية التي تم إنجازها بالفعل؛ منها النصية ومنها الصوتية ومنها المختلَطة.

ومن الجدير بالذكر في هذا المقام أن إعداد وعنونة المواد اللغوية الخام - سواء كانت نصية أم لا، والتي تسمى اصطلاحًا «مُدَوَّنات» (Corpora)لإنتاج الموارِد اللغوية قد صار منذ بداية القرن الحادي والعشرين الميلادي نشاطًا ذا حجم كبير في أبحاث وصناعة تطبيقات معالجة اللغات الحية بمختلَف أصنافها، وقد يتخطى حجم هذا النشاط أحيانًا الخمسين في المئة ضمن كافة أنشطة معالجة اللغات الحية من معالَجات رياضية وتصميم وبناء البرمجيات اللازمة لتنفيذها ... إلخ، وهو ما أدركه المجتمع العالمي المشتغل بهذا المجال فقام بتخصيص أكبر مؤتمراته للموارِد اللغوية وبمختلف أنواع هذه الموارِد ويُعْرَف هذا المؤتمر باسم LRECويُعْقَد بوتيرة منتظمة مرةً كل سنتين منذ عام ١٩٩٨م ويتراوح عدد حضوره في كلٍّ من دوراته ما بين الألف والألفين من المتخصصين كما يتجاوز عدد مساراته العشرين مسارًا كي تغطي مختلف أنواع وأنماط الموارِد اللغوية http: / /www.lrec-conf.org.

إعداد المدَوَّنات اللغوية وبناء الموارِد اللغوية بمختلف أنواعها هو أحد أكبر الأنشطة في مجال أبحاث وصناعة تطبيقات معالجة اللغات الحية. وبسبب ضخامة حجم المادة وارتفاع دقة العنونة المطلوبَيْن لكلٍّ من هذه الموارِد اللغوية، فضلاً عن التنوع الواسع لأصناف هذه الموارِد، إضافةً إلى قلة عدد الموارِد اللغوية التي جرى إنجازها بالفعل في حقل اللغة العربية، فإن هذا النشاط يستدعي أن تنخرط فيه أعداد كبيرة من اللغويين العرب المؤهَّلين والملِمِّين بطبيعة تقنيات المعالجة الحاسوبية لِلُّغة مما نَعُدُّه أحد الأهداف التي نرجو أن تتحقق من وراء تأليف هذا الكتاب.

١٠- أدوات وبرمجيات ومعايير

يتعذر بشكل واقعي إجراء عملية بناء الموارِد اللغوية على وجه العموم خارجَ إطار تنظيمي يضبط اطِّراد واتساق عنونة المدوَّنات اللغوية (مثل التحليل الصرفي لكلمات المدونة) من مُعَنْوِنٍ لآخر، ويدعم في الوقت نفسه كفاءة العنونة ودقتها، ويتيح كذلك مراجعتها ومراقبة نسبة الإنجاز ... وغير ذلك من وظائف. وغالبًا ما يُصاغُ كل إطار من مثل هذه الأُطُر على هيئة برنامج حاسوبي ذي بيئة تفاعلية يُسَمَّى «أداة عَنْوَنة» (Annotation Tool).

ويعرض الشكل رقم (٣-٥) أدناه لقطة أثناء تشغيل إحدى هذه الأدوات التي تمكن اللغويين من إدارة وعنونة المدونات النصية العربية من النواحي الصوتية (الفونولوجية) والصرفية والدلالية وذلك بأسلوب تفاعلي عبر واجهة استخدام رسومية مفصلة تتيح لهم مراجعة عملهم هذا ومراقبة نسب الإنجاز فيه.

	الحَلّ ١	الحَلّ ٢	الحَلّ ٣	الحَلّ ٤	الحَلّ ٥
الكَلِمة	مَرّ	مَرّ	مَرّ	مَرّ	مَرّ
نَوْعُها	مُصرّفة منتظِمة	مُصرّفة منتظِمة	مُصرّفة منتظِمة	مُصرّفة منتظِمة	مُصرّفة منتظِمة
السَّابِق					
الجِذْر	مرر	مرر	مرر	مرو	مري
الوَزْن	فَلّ	فَلّ	فَلّ	فَعّل	فَعّل
اللَّاحِق					
الأَرْجَحيَّة	٪٩٤٫٤٠٣	٪٥٫٥٩٦	٪٠٫٠٠١	٪٠	٪٠

الشَّكل ٣-٥: لقطة من أداة عنونة النص العربي صوتيًّا وصرفيًّا ودلاليًّا؛ «فَصيح»[١].

وفي حين تجعل طبيعةُ بعض الموارِد اللغوية أدواتِ العنونة المستخدَمةَ لإنتاجها غيرَ مُعتمِدةٍ في وظائفها على لغة محتوى هذه الموارِد (مثل بعض أنواع الموارِد اللغوية للكلام المنطوق) فتكون تبعًا لذلك مثلُ هذه الأدواتِ قابلةً لإعادة الاستخدام في أغلب الأحيان بين لغة وأخرى، فإن طبيعةَ بعض الموارِد اللغوية الأخرى (وخصوصًا تلك النصية، أو تلك التي يشكل النص ركنًا أساسيًّا فيها) تحتم إنشاء أدوات عنونة جديدة تراعي خصوصيات لغة محتوى هذه الموارِد، وتقع أدوات عنونة المدَوَّنات النصية العربية بمختلف أنواعها ضمن الفئة الثانية، وبنظرة عامة على الحالة الراهنة للأدوات العربية من هذه الفئة الثانية تتضح لنا المساحات التالية التي تتسع لإسهامات اللغويين العرب:

١- يمكن الرجوع إلى تفاصيل هذه الأداة التفاعلية في الورقة المشار إليها برقم [١٦] في قائمة المراجع نهاية هذا الكتاب.

- بالقياس إلى الكثير الذي يُرتقَب إنجازُه والقيام على تحديثه من مختلف الموارِد اللغوية العربية النصية (أو تلك التي يُعتبَر النص العربي ركنًا فيها) يُعَدُّ ما جرى بناؤه من أدوات العنونة العربية اللازمة لتلك الموارِد قليلاً. وإلى جانب الجهود العلمية والهندسية والبرمجية المطلوبة لبناء الأدوات الناقصة، فإن هناك كذلك حاجة إلى إسهام من ناحية اللغويين العرب وخصوصًا فيما يتعلق بتصميم واختبار واجهات الاستخدام وتفاعُلِيَّتِها.
- وحتى بالنظر إلى أدوات العنونة العربية التي تم بناؤها بالفعل فإن استخدامها لا يزال غير شائع إلا على نطاق محدود، ومرجع ذلك في كثير من الأحيان هو أن ملكيتها تعود إلى الشركات التي طورتها، ولذلك فإن هناك حاجة للعمل على التوعية بالأدوات الجيدة بين أوساط اللغويين وخصوصًا تلك المتاحة للاستخدام الحر، فضلاً عن اقتراح أنماط مبتكرة من تعويض الشركات لإقناعها بإتاحة أفضل ما طورته من أدوات العنونة العربية للاستخدام الحر أو مقابل رسوم زهيدة، وهو ما يحتاج إلى جمعيات/ تنسيقيات/ تكتلات من المشتغلين بأبحاث وصناعة تطبيقات المعالجة الحاسوبية للغة العربية بمن فيهم اللغويون.
- إجراء رصد دوري شامل لكل ما يُطْرَح من أدوات العنونة العربية مع نقد وتقييم أداء كل منها ومن ثَمَّ ترشيح الأفضل من بينها لإنجاز كل نوع من الموارِد اللغوية العربية، فضلاً عن إجراء رصد أوسع لأدوات العنونة غير العربية ذات البرمجيات مفتوحة المصدر مما قد يمكن تطويعها بجهد معقول لإنتاج الموارِد اللغوية العربية، ومثل هذه الأرصاد والتقييمات تحتاج أيضًا لجهود منظَّمة من جمعيات/ تنسيقيات/ تكتلات من المشتغلين بأبحاث وصناعة تطبيقات المعالجة الحاسوبية للغة العربية بمن فيهم اللغويون.

أما بالنظر إلى الوضع الراهن لبرمجيات التحليل الصرفي العربي بشِقَّيْها؛ الأول المختص باستخلاص كل التحليلات الصرفية المُمْكِنة لأية كلمة عربية والثاني المختص بفك الالتباس الصرفي، فإن تاريخ العمل على هذه المسألة من بحث وتطوير والذي يمتد منذ سبعينات القرن العشرين الميلادي قد أثمر عددًا كافيًا من حزم البرمجيات الناضجة بمستوى تطبيقي صناعي وذلك مع نهاية تسعينات القرن العشرين الميلادي، حيث وصل هامش الخطأ في الشق الأول إلى ما دون الاثنين في الألف كما وصل هامش الخطأ في الشق الثاني إلى ما دون الخمسة في المئة بعد التدريب على ذخيرة صرفية عربية متوازنة

التغطية للفضاء اللغوي ذات حجم ثلاثة ملايين كلمة [1] وذات جودة عنونة رفيعة لا يزيد هامش الخطأ فيها عن خمسة في الألف، وفي حدود هوامش خطأ محدودة كهذه أمكن بناء تطبيقات صناعية ترتكز على التحليل الصرفي العربي الآلي وذات أداء مقبول.

ويجدر بنا الإشارة في هذا المقام قبل أن نختتم هذا الفصل إلى بعض الأعمال التي اشتهرت وشاعت بين الباحثين في مجال المعالجة الآلية لِلُّغة العربية؛ فلدينا المحلل الصرفي العربي الذي اشتهر باسم مصممه الأساسي «تِيمْ باكْوُولْتَرْ» Tim Buckwalter (http://www.qamus.org/morphology.htm) الذي يمكن الحصول عليه من خلال «مُؤَسَّسة Linguistic Data Consortium (LDC)» في جامعة «بِنْسِلْفانْيا» بالولايات المتحدة الأمريكية أثناء عمله فيها وذلك عبر http:// www.ldc.upenn.edu/catalog/catalogEntry.jsp?cataloId=LDC2002L49 والنسخة الشائعة منه هي الإصدارة ١,٠ التي ظهرت في نوفمبر عام ٢٠٠٢م ويمكن الحصول عليها بالمجان للمشتركين في الجمعية أو بمقابل رمزي لغير المشتركين فيها.

كما أن هناك حزمة الأدوات البرمجية المعروفة باسم «مَدَى وطُوقَانْ» (MADA+ TOKAN) والتي تعمل على معالجة النص العربي الخام لاستخراج أنسب المعلومات الصرفية والمعجمية والصوتية والأنواع الكلامية لكل كلمة بمراعاة السياق المحيط بها، وقد تم تطوير هذه الحزمة في جامعة «كُولُومْبِيا» بالولايات المتحدة الأمريكية ويمكن الاطلاع على تفاصيلها في الورقة المشار إليها برقم [١٩] فى قائمة المراجع نهاية هذا

١- وقد ينخفض هامش الخطأ في فك الالتباس الصرفي الآلي إلى ما دون الثلاثة في المئة مع ذخيرة صرفية بنفس الحجم ولكنها مركزة في حيز ضيق من الفضاء اللغوي («النصوص القانونية المعاصرة في مصر» على سبيل المثال)، كما يُعتَقَد أنه دون الاستعانة بتحليلات صحيحة من طبقات لغوية أعلى (كالدلالة والنحو على سبيل المثال) يظل هناك حدٌّ أدنى لهذا النوع من هوامش الخطأ لا يقل عن الثلاثة في المئة مهما زاد حجم الذخيرة اللغوية الصرفية التي جرى تدريب آلية فك الالتباس عليها.

الكتاب.

كما تعالج حزمة الأدوات البرمجية المعروفة باسم «أَمِيرَة» (AMIRA) كلمات النص العربي لتعيين لواصقها وتحديد أنواعها الكلامية وتقسيم النص إلى أقصر عبارات غير متداخلة في شجرة تحليله الإعرابي، ويمكن الاطلاع على تفاصيلها في الورقة المشار إليها برقم [١٨] في قائمة المراجع نهاية هذا الكتاب.

ومع عدم الحاجة للمزيد من الأبحاث في مسألة التحليل الصرفي العربي الآلي، فإنه لاتزال هناك ضرورة للعمل من أجل تحقيق الأهداف التالية:

- الاتفاق على معيار موحَّد لمدْخَلات ومُخْرَجات برمجيات التحليل الصرفي، مما يسمح بتعظيم فرص إعادة استخدامها في التطبيقات الصناعية، أو ودمجها في آليات أوسع للمعالجة الحاسوبية لِلُّغة العربية.
- الاتفاق على معيار موحَّد وعلى عينة اختبار موحَّدة لأداء برمجيات التحليل الصرفي العربي مما يُمَكِّن من المقارنة العادلة بَيْنَها.
- رصد كل ما يمكن الوصول إليه من برمجيات التحليل الصرفي العربي - وخصوصًا تلك المتاحة بالفعل للاستخدام الحر - فضلاً عن اقتراح أنماط مبتكرة من تعويض الشركات لإقناعها بإتاحة أفضل ما طورته من المحللات الصرفية العربية الآلية للاستخدام الحر - أو مقابل رسوم زهيدة - بل وحتى طرحِها كبرمجيات مفتوحة المصدر.
- كل ما سبق من أعمال يحتاج أيضًا لجهود منظَّمة من جمعيات/ تنسيقيات/ تكتلات من المشتغلين بأبحاث وصناعة تطبيقات المعالجة الحاسوبية للغة العربية بمن فيهم اللغويون.

ببليوجرافيا مرجعيَّة

١. ابن فارِس (أَبُو الحُسَيْنِ أَحْمَد): مَقاييس اللُّغَة، تحقيق: عَبْدالسَّلام هارون، ط٢، ١٩٦٩م.

٢. ابن مَنظُور (أبو الفَضل جمال الدِّين): لِسانُ العَرَب، دارُ فاضل، بَيْروت.

٣. الدَّحْداح (أَنْطُوان): مُعْجَمُ قَواعِدِ العَرَبِيَّةِ العالَمِيَّة، مَكْتَبة لُبْنان، ط١، ١٩٩٠م.

٤. الرَّاجِحيّ (عَبدُه): التَّطبيقُ الصَّرْفيُّ، دارُ المعْرِفةِ الجامِعِيَّةِ، الإسْكَنْدريَّةُ، ١٩٩٣م.

٥. رِضا (أحمد): مُعْجَم مَتْن اللُّغَة، أَحْمَد رِضَا، مَكْتَبَة الحَيَاة، بيْروت، ١٩٦٠م.
٦. الزَّبيديّ (مُحَمَّدُ مُرْتَضى): تاجُ العَروسِ مِنْ جَواهِرِ القاموس، طَبْعة وَزارةِ الإعلامِ الكُوَيْتيَّة.
٧. العاني (سَلمان حَسَن): فُونولوجيا العربيَّة، - ترجمة ياسِر المَلَّاح، دار النَّادي الأَدَبِيِّ بجدَّة - المَمْلَكة العَرَبيَّة السَّعوديَّة، ١٩٨٣م.
٨. عَبْد الفَتَّاح (إبراهيم أَحْمَد): القاموسُ الطَّويلُ لِلْقُرْآنِ الكَريمِ، مَجْمَعُ البُحوثِ الإسلاميَّةِ، ١٩٨٣م.
٩. عُضَيْمة (مُحَمَّد عَبْدُ الخالِق): المغْنِي في تصريفِ الأَفْعال، دارُ الحَديث.
١٠. الفارابِيّ (أَبُو إِبْراهيمَ إِسْحاقَ ابْنُ إبراهيمَ): دِيوان الأدب، تحقيق: أَحْمَد مختار عُمَر، ١٩٧٤م.
١١. فَيَّاض (سُلَيمان): الحُقولُ الدَّلاليَّةُ الصَّرْفيَّةُ لِلأَفْعالِ العَرَبيَّة، دار المرِّيخ بالرِّياض، ١٩٩٠م.
١٢. مَجْمَعُ اللُّغَةِ العَرَبيَّةِ بالقاهِرةِ، المعْجَمُ الوَسيطُ، ط٣، ١٩٨٥م.
١٣. مَجْمَعُ اللُّغَةِ العَرَبيَّةِ بالقاهِرةِ، ما تَمَّ مِنْ (المعْجَم الكَبير)، ط١، حَرْفُ الأَلِف ١٩٧٠، حَرْفُ الباء ١٩٨١، حَرْفا التَّاء والثَّاء ١٩٩٢، حَرْفُ الجيم ٢٠٠٠، حَرْفُ الحاء ٢٠٠٠، حَرْفُ الخاء ٢٠٠٤، حرف الدال ٢٠٠٦، حرف الذال ٢٠٠٨م، حرف الرَّاء ٢٠١٥م.
١٤. مَجْمَعُ اللُّغَةِ العَرَبيَّةِ بالقاهِرةِ، مُعْجَمُ أَلْفاظِ القُرْآنِ الكَريمِ، طَبْعةٌ مُنَقَّحةٌ، ١٩٩٠م.
15. Attia, M., A Large-Scale Computational Processor of the Arabic Morphology, and Applications, M.Sc. thesis, Dept. of Computer Engi-neering, Faculty of Engineering, Cairo University, 2000.
16. Attia, M., Rashwan, M., Al-Badrashiny, M., Fassieh©; a Semi-Automatic Visual Interactive Tool for the Morphological, PoS-Tags, Phonetic, and Semantic Annotation of the Arabic Text, IEEE Transactions

on Audio, Speech, and Language Processing (TASLP) http://www.SignalProcessingSociety.org/Publications/Periodicals/TASLP: Special Issue on Processing Morphologically Rich Languages, Vol. 17 - Issue 5; pp. 916-925 http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?isnumber=5067414&arnumber=5075778&count=21&index=6, July 2009.

17. Cavalli-Sforza, V., Soudi, A., Mitamura, T., Arabic Morphology Gen-eration Using a Concatenative Strategy, ACM Int'l Conference Pro-ceeding Series; Proceedings of the first conference on North Ameri-can chapter of the Association for Computational Linguistics (NAACL), 2000.

18. Diab, M., Second Generation AMIRA Tools for Arabic Processing: Fast and Robust Tokenization, POS tagging, and Base Phrase Chunking, The Proceedings of the 2nd Intl. Conference on Arabic Language Resources and Tools, ISBN: 2-9517408-5-9, Cairo – Egypt, http://www.medar.info/conference_all/2009/index.php, Apr. 2009.

19. Habash, N., Rambow, O., Roth, R., MADA+TOKAN: A Toolkit for Ara-bic Tokenization, Diacritization, Morphological Disambiguation, POS Tagging, Stemming and Lemmatization, The Proceedings of the 2nd Intl. Conference on Arabic Language Resources and Tools, ISBN: 2-9517408-5-9, Cairo – Egypt: http://www.medar.info/ conference_all/ 2009/ index.php, Apr. 2009.

20. Schütze, H., Manning, C. D., Foundations of Statistical Natural Language Processing, MIT Press, 2000.

21. Vapnik, V. N., Statistical Learning Theory, John Wiley & Sons Inc., 1998.

22. Yaseen, M., et al., Building Annotated Written and Spoken Arabic LR's in NEMLAR Project, LREC2006 conference http://www.lrec-conf.org/lrec2006, Genoa-Italy, May 2006.

الفصل الرَّابع

التَّحليل التَّركيبيّ

د. سامح الأنصاري

١- تقديم.

٢- علاقة منظومة النحو بمنظومتي الصرف والمعجم.

٣- مدخل إلى تركيب الجملة في علم اللغة.

٤- مدخل إلى معالجة تركيب الجملة آليا.

٥- اتجاهات بناء المحللات النحوية.

٦- دقة المحللات النحوية.

٧- أمثلة للمحللات النحوية.

٨- بعض النقاط البحثية الهامة من أجل دعم خطة طريق لمعالجة التركيب في اللغة العربية.

١ - تقديم

إن تناول قضية التحليل النحوي والمعالجة الآلية لتركيب الجملة في اللغة العربية موضوع متشعب الأبعاد والجوانب نتيجة لخصوصية طبيعة اللغة العربية، حيث إن منظومة النحو تندرج فيها العديد من النظريات النحوية القديمة والحديثة، بالإضافة إلى ارتباطها ارتباطًا وثيقًا بمنظومتي الصرف والمعجم، ولا مناص من مناقشة التفاعل بينهم؛ التفاعل الذي يصل إلى حد الاندماج الكامل.

وتعتبر عملية التحليل النحوي الآلي من ركائز المعالجة الآلية للغات الطبيعية، حيث يجري فيها تحديد بنية الجملة من حيث هيكلية مكوناتها ووظائف عناصرها، وإيجاد قالبها النحوي اعتمادًا على القواعد النحوية الأساسية من حيث تقسيم الجملة وتحديد مكوناتها وتقسيم كلماتها لإيجاد العلاقات النحوية فيما بينها. ويعد المحلل النحوي الآلي مقومًا أساسيًا لا غنى عنه، حيث تعتمد عليه العديد من تطبيقات معالجة اللغة مثل التطبيقات الخاصة بتحليل مضمون النصوص وفهمها آليًا وعمليات الاستخلاص والتلخيص والفهرسة الآلية، ونظم الترجمة الآلية، وتطبيقات اكتشاف الأخطاء الهجائية والنحوية آليًا.

وقبل التعرض لآليات التحليل النحوي، لا بد من النظر - بشكل عام ودون التقيد باتجاه معين - إلى مفهوم القواعد النحوية وكيفية التعبير عنها وتمثيلها وآليات تحليل مكوناتها، فنحو اللغة كان أحد موضوعات البحث منذ وقت طويل في أبحاث علوم اللغة والآداب. ونظرًا لطبيعة النحو التي تميل إلى التجريد والاختزال، فقد أصبح النحو - كما ذكر الدُّكتور نبيل علي - «خط الالتقاء بين اللسانيات والرياضيات (علي، ١٩٨٨). ومؤخرًا قدم مجال علوم الحاسب رؤية جديدة للتمثيل النحوي؛ وهي أن التركيب النحوي يمكن تعريفه على أنه نتاج عدد من العمليات الحسابية والتي يمكن من خلالها صياغة قواعد النحو في صورة قواعد رياضية، يمكن من خلالها توليد عدد لا نهائي من التراكيب اللغوية المسموح بها في اللغة.

فيهدف هذا الفصل إلى تقديم بعض المفاهيم الأساسية المتعلقة بتركيب الجملة لترسيخ مفهوم التحليل النحوي، وتوضيح أهدافه. فسيسعى هذا الفصل أيضًا إلى توضيح مفهوم القواعد والقدرة التوليدية للقواعد وقياسها وإلقاء الضوء على الفارق

بين بعض الاتجاهات المختلفة للتحليل النحوي، واستعراض التحديات التي تواجهه في اللغة العربية بشكل خاص. ثم يتم رصد بعض المحاولات السابقة لبناء المحللات النحوية الآلية. ويلي ذلك عرض عام لتقديم المحلل النحوي التفاعلي (Interactive Analyzer -IAN)كيفية عمله. وينتهي هذا الفصل بالتعرض للعديد من النقاط البحثية الهامة من أجل دعم خطة طريق لمعالجة التركيب في اللغة العربية.

٢- علاقة منظومة النحو بمنظومتي الصرف والمعجم

تعتبر اللغة العربية من أعقد اللغات في عائلتها ومن أغناها صوتًا وصرفًا ومعجمًا ونحوًا. ومما لا شك فيه أن النظر إلى طبيعة منظومة النحو العربي يستدعي التطرق إلى طبيعة منظومة الصرف. إذ أن العلاقة العضوية بين الصرف والنحو قضية تفرض نفسها على كل من الساحة اللغوية الوصفية وساحة المعالجة الآلية للغة على حدٍ سواء. وقد تصل عملية التفاعل بين المنظومتين إلى حد الاندماج الكامل إلى الدرجة التي يصعب فيها تحديد أين تلتقيان وأين تفترقان. فعلاقة النحو بالصرف نقطة تحول كبيرة في النظرية النحوية الحديثة، فقد قام تشومسكي[١] بإدراج «الصرف بأكمله في منظومة النحو في نظريته القياسية؛ حيث نظر تشومسكي إلى عمليات التكوين الصرفي للكلمات بصفتها عمليات تحويل نحوية» (علي، ١٩٨٨). فالجمل ليست مجرد تتابع للكلمات، بل هي هيكلية تتعدد مستوياتها وتتداخل. ونعني بذلك مثلاً أنه من غير الممكن تصور إمكانية تحديد اللواحق الصرفية دون معرفة الموقع الإعرابي للكلمة فعلى سبيل المثال لا يمكن معرفة أن اللاحقة الصرفية المسئولة عن تعيين الجمع المذكر لكلمة) معلم) هي (ـون) أم (ـين) دون النظر إلى الموقع الإعرابي للكلمة داخل الجملة كما في (١)، (٢):

حضر المعلمـ(ـون) إلى الاجتماع. (١)

كافأ المدير المعلمـ(ـين) القدامى. (٢)

ويمكن أيضًا أن يكون للصيغة الصرفية الواحدة أكثر من وظيفة نحوية. فإذا نظرنا إلى اللاحقة الصرفية (ـت) الملحقة بالفعل (كنت) فقد يشعر القارئ بالغموض

١- درس نعوم تشومسكي في مجالات متعددة وهي مجالات اللِّسانيَّات والرياضيات والفلسفة وهو مؤسس النظرية التوليدية التحويلية.

وخصوصًا مع عدم وجود الحركات الإعرابية في الجمل (٣)، (٤)، (٥):

كنـ(ـت) أعمل على الورقة البحثية أمس. (٣)

كنـ(ـت) تعمل على الورقة البحثية أمس. (٤)

كنـ(ـت) تعملين على الورقة البحثية أمس. (٥)

فيصعب تحديد نوع التاء الملحقة بكل فعل - هل هي تاء الفاعل للمتكلم (كنتُ) كما في (٣) أم تاء الفاعل للمخاطب المذكر (كنتَ) كما في (٤) أم تاء الفاعل للمخاطب المؤنث (كنتِ) كما في (٥) - دون معرفة الفاعل في البنية التركيبية لكل جملة. وعلى ذلك فإنه ليس من المتصور أن تتم دراسة التركيب النحوي للجملة بمعزل عن التركيب الصرفي للكلمات.

كما أن علاقة النحو بالمعجم علاقة محورية في منظومة اللغة لما يزودنا به من أبنية الأسماء والأفعال وصيغ الاشتقاق. فلم يعد المعجم كيانًا مستقلاً، بل هو مصدر لإمداد النحو بالمفردات التي تغذي المنظومة النحوية ولتحديد أنواع هذه المفردات والفئات التي تنتمي إليها. فعلى سبيل المثال جملة (حضر الطالب متأخرًا إلى المدرسة) نجد أنه لا يمكن التطرق لتحليل الجملة ووصف تركيبها دون تحديد الفئات التي تنتمي إليها كل كلمة وتعيين سماتها الصرفية. وهنا يأتي دور المعجم المسئول عن هذه الوظيفة ليمدنا بمثل هذه المعلومات: حضر: فعل)(Verb (V)، ال: أداة تعريف ((Determiner(DET))، طالب: اسم ((Noun (N)، متأخرًا: حال (Adverb (ADV))، إلى: حرف) (Preposition(P)، مدرسة : اسم (N). ويجب أن ننوه أن علاقة النحو العربي بالمعجم تتم - في معظمها - من خلال الصرف وذلك نظرًا للثراء الاشتقاقي الشديد للصرف العربي[1]. والخلاصة أن السمات الصرفية النحوية هي إحدى الركائز الأساسية التي لا بُدَّ منها لصياغة النحو العربي وللمعالجة الآلية للغة.

١ - «الجملة العربية ما هي إلا نتاج التفاعل بين إسقاطات كلماتها النحوية والمعجمية».

٣- مدخل إلى تركيب الجملة في علم اللغة

تتكون اللغة من مجموعة من الحروف التي تتكون منها الكلمات والتي بدورها مسئولة عن تكوين الجمل، وتختلف كل لغة عن الأخرى في العديد من الخصائص اللغوية ومن أهم هذه الخصائص ترتيب الكلمات. فلكل لغة تركيب نحوي أي توزيع منتظم غير عشوائي للعناصر المكونة لكل جملة؛ ويتم ذلك التوزيع من خلال مجموعة من القواعد؛ مسئولة عن تحديد هيكل الجملة وترتيب عناصرها وتوصيف وظيفة كل عنصر. فعلى سبيل المثال في اللغة الانجليزية يأتي الفاعل قبل الفعل بينما يأتي المفعول به بعد الفعل ويسمى هذا التركيب (Subject –Verb –Object (SVO)) ولكن الوضع ليس كذلك في كثير من اللغات الأخرى فقد يأتي الفاعل والمفعول قبل الفعل. لذلك فإن أحد الوظائف الهامة للقواعد هي توصيف كيفية التعبير عن تركيب الجمل. ويُعرِّف تشومسكي قواعد اللغة[١] على أنها تتكون من ثلاث مكونات أساسية وهي: القواعد النحوية ويقوم هذا المكون بتحديد القواعد التركيبية المحدِدة لتركيب الجمل تركيبا سليما في اللغة. والقواعد الدلالية والتي تحدد كيفية تفسير معنى الجمل. والقواعد الفونولوجية وهو المكون المحدِد للنظام الصوتي الحاكم للجانب المنطوق من اللغة. ولا بد هنا أن ننوه إلى أن فقدان المكون المورفولوجي من ضمن تلك المكونات يرجع إلى أن تشومسكي يتبنى الفكر الذي يزكي كون المكون المورفولوجي يتداخل ضمن المكونين النحوي والفونولوجي (Katamba, 1993).

وبالنظر إلى مُكون القواعد النحوية- قضيتنا الأساسية في هذا الفصل- فإن هذا المكون يجب أن يؤدي وظيفتين جوهريتين وهما؛ أولا: القدرة على التوليد والتحليل الصحيحين لكل الجمل الصحيحة فقط، ثانيا: تعيين البنية النحوية المناسبة للجمل المعنية بالتحليل مع الأخذ في الاعتبار الملكة اللغوية [٢] لابن اللغة وإدراكه للعلاقات النحوية بين الكلمات في هيكل الجملة .وقبل الخوض في تفاصيل هاتين الوظيفتين لا بُدَّ من إلقاء الضوء على مفهوم (التركيب النحوي) للجملة وكيفية تمثيل هذا التركيب

١- كلمة قواعد تعني وصف اللغة من حيث هي واقع قائم بذاته وصفا موضوعيا وبإمكانها توليد جمل اللغة أي وصف جمل اللغة بصورة بينة وجلية فمع نشوء اللِّسانيَّات وتطورها لم تعد كلمة قواعد تعني القوانين التي يجب اتباعها للتعبير والكتابة بحسب الأصول.

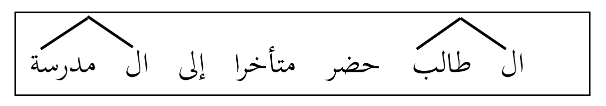
٢- الملكة اللغوية: هي جزء من كفاية الإنسان، أي جزء من معرفته الضمنية بقواعد اللغة.

والقواعد المنظمة لهذا التمثيل. ويعد التمثيل الهرمي من أنجح طرق التمثيل التركيبي للجمل الطبيعية وذلك لأنها تحاكي الملكة اللغوية لابن اللغة الذي لديه معرفة بديهية بتركيب الجمل في لغته؛ فقدرته على بناء مجموعات صحيحة تبدأ من التعرف على مجموعة معينة من الأصوات يمكن أن تكون كلمة صحيحة قائمة بذاتها منفصلة عن تتابع آخر، ويستطيع ابن اللغة أن يدرك أن تتابع الكلمات يمكن أن يشكل مكون أو مركب أكبر وهو الوحدة المكونة للجملة؛ فعلى سبيل المثال يمكن تمثيل تيار أصوات الجملة في (٦) بدون مسافات. وهنا نجد أن أي ابن للغة العربية لديه ملكة لغوية تجعله قادرا على تمييز وتقسيم تلك التتابعات الصوتية الموجودة في (٦) إلى تلك الموجودة في (٧) بشكل غير واع:

الطالبحضرمتأخرا (٦)

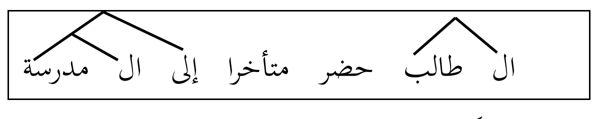
الطالب حضر متأخرا إلى المدرسة(١). (٧)

بالإضافة إلى قدرة ابن اللغة على تقسيم تتابعات الأصوات إلى كلمات فإن هذه القدرة تمتد لبيان الرابط التركيبي بين الوحدات، فيستطيع أن يتعرف على العلاقة الوثيقة بين «ال» و «طالب»؛ وهي من نفس نوع العلاقة التي تربط بين «ال» و«مدرسة» حيث ترتبط «ال» بـ «مدرسة» وليس بالوحدة التي تسبقها «إلى» كما هو مبين في الشكل (٤-١).



الشَّكل ٤-١: بيان الرابط التركيبي بين الكلمات.

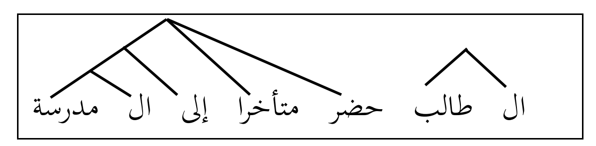
وبالمثل يعرف ابن اللغة بشكل غير واع أن «المدرسة» مرتبطة بـ «إلى» لتمثل وحدة أكبر كما هي ممثلة في الشكل (٤-٢).



الشَّكل ٤-٢: الرابط بين «المدرسة» و «إلى».

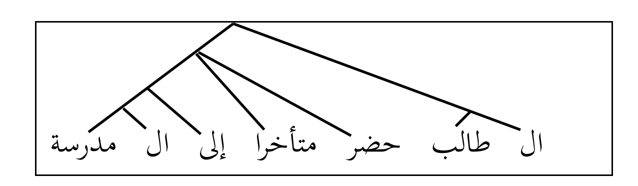
١- وفقًا لمعياريَّة اللُّغَة العربيَّة، يُفتَرَضُ أن يُقَدَّمَ الفِعلُ على الاسم في هذا النَّمَط التَّركيبيّ. ومع هذا يجوزُ تقديم الاسم على الفعل، كما في قَولِهِ تعالى ﴿والله يَعصِمُكَ مِنَ النَّاس﴾ [المائدة/ ٦٧]. وقد آثَرَ المؤَلِّفُ التَّمثيلَ بهذا النَّمط لتيسير فهم القاعدة على اللُّغات الإنسانِيَّة عُمومًا، حيثُ يشيعُ تقديمُ الاسم على الفِعل.

وتمتد هذه القدرة لتربط «حضر» و «متأخرا» و «إلى المدرسة» كوحدة واحدة كما في الشكل (٤-٣).



الشَّكل ٤-٣: ربط «حضر» و«متأخرا» و«إلى المدرسة» كوحدة واحدة.

وأخيرًا تتابع الكلمات «الطالب حضر متأخرا إلى المدرسة» يمثل وحدة واحدة قائمة بذاتها كما هي ممثلة في الشكل (٤-٤) والذي هو عبارة عن تمثيل تخطيطي لما يدور في عقل ابن اللغة عن العلاقات النحوية بين سلسلة الكلمات المتتابعة ويبرز قدرة ابن اللغة على تحديد الوحدات التي تتكون منها الجملة في (٧).



الشَّكل ٤-٤: تمثيل تخطيطي للوحدات التي تتكون منها الجملة (٧).

كذلك تمتد قدرة ابن اللغة ليستطيع التعرف على الوحدات التي لها نفس الفئة وذلك من خلال ملكته اللغوية الذي يمتلكها؛ فعلى سبيل المثال فإنه يستطيع أن يدرك بديهيا أن «طالب» و «مدرسة» وحدتان تنتميان لنفس الفئة والتي تختلف تماما عن فئة الكلمة «حضر» وذلك لأن كلا من «طالب» و «مدرسة» اسم؛ ولكن «حضر» فِعل. ويستطيع أيضا ابن اللغة أن يدرك أن «متأخرا» لها فئة مختلفة وهي حال والتي تختلف عن فئة «إلى» وهي حرف وكذلك «ال» التي تسبق «طالب» هي نفسها تلك التي تسبق «مدرسة» ويطلق عليهما أداة تعريف. وتعرف هذه الفئات بالفئات المعجمية.

ويمكن كذلك لابن اللغة أن يدرك فئة الوحدات الأكبر ويميزها عن غيرها من الوحدات داخل الجملة؛ فهو يستطيع -على سبيل المثال - أن يدرك أن «الطالب» و«المدرسة» وحدتان لهما نفس الفئة وهي مركب اسمي ((Noun phrase (NP))، ويطلق عليه ذلك لأن فئة رأسه (اسم)؛ وأن الوحدة «إلى المدرسة» تُشكل مركبًا حرفيًّا

وسمي حرفيًّا (Prepositional phrase (PP))، كذلك لأن فئة رأسه (حرف)؛ وأن «حضر متأخرا إلى المدرسة» تُمَثِّلُ مُرَكَّبًا فعليًّا ((Verbal phrase (VP))، وسمي كذلك لأن فئة رأسه (فعل)، وأخيرا، أنَّ التتابع «الطالب حضر متأخرا إلى المدرسة» يُمَثِّلُ الوحدة الأكبر التي يطلق عليها (الجملة) ((Sentence (S))؛ وتعرف هذه الفئات بالفئات العبارية (Phrasal Categories).

وهذا الفكر يطابق ما عرفه اللغويون بالوسم العباري (Phrase Marker) الذي يعرف اختصارًا بـ (P-marker)؛ وهو يمثل الهيكل التركيبي للجملة في شكل تتابع هرمي من الكلمات هذه الكلمات تشكل مركب (Phrase) وتتابع المركبات يكون جملة. ويتم توصيف هذا الهيكل التركيبي في شكل تمثيل شجري، ومثال على ذلك جملة (حضر الطالب متأخرا إلى المدرسة): يتم أولا تعيين الفئة التي تنتمي إليها كل كلمة في الجملة وذلك يتم من خلال المعجم. فـ «الطالب» تتكون من أداة التعريف «ال» والاسم «طالب» والفعل «حضر» والحال «متأخرا» والحرف «إلى» أما «المدرسة» فتتكون من أداة التعريف «ال» والاسم «مدرسة» كما في الشكل (٤-٥).

N	DET	P	ADV	V	N	DET
مدرسة	ال	إلى	متأخرا	حضر	طالب	ال

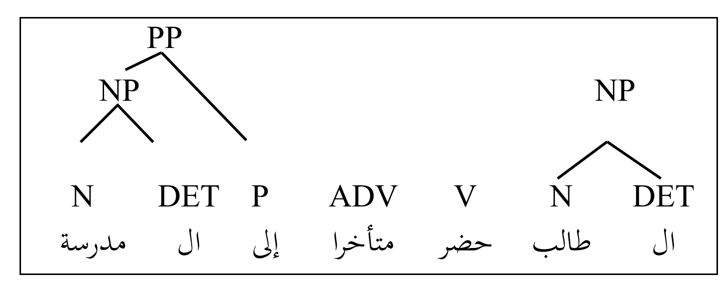
الشَّكل ٤-٥: فئات الكلمات التي تتكون منها الجملة (٧).

ويتم بعد ذلك جمع الكلمات التي تشكل مركبات، فنجد أن أداة التعريف «ال» تكون مع الاسم «طالب» مركبًا اسميًّا وأداة التعريف الأخرى «ال» تكون مع الاسم «مدرسة» مركبًا اسميًّا كما في الشكل (٤-٦).

NP					NP	
N	DET	P	ADV	V	N	DET
مدرسة	ال	إلى	متأخرا	حضر	طالب	ال

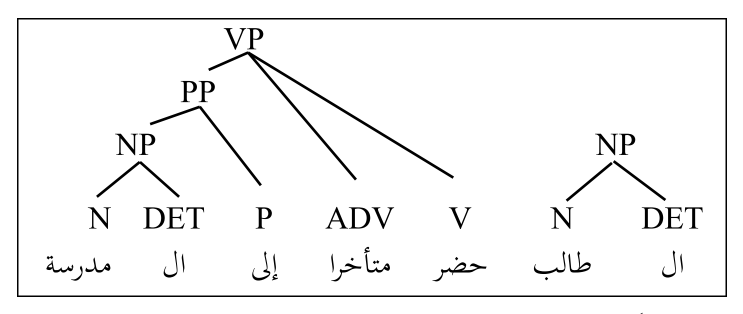
الشَّكل ٤-٦: ربط الكلمات التي تشكل مركب اسمي.

وأن المركب الاسمي «المدرسة» يكون مع الحرف «إلى» مركبًا حرفيًّا (PP) كما في الشكل (٤-٧).



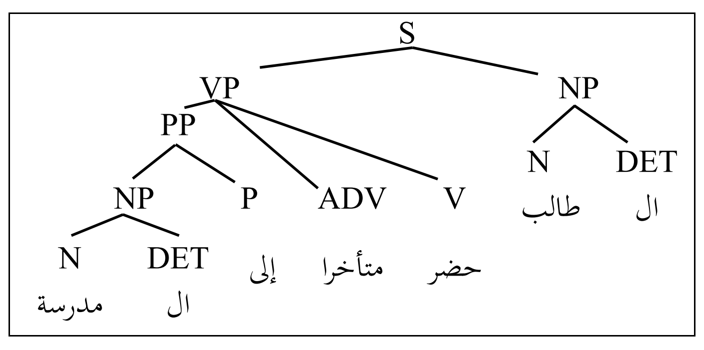
الشَّكل ٤-٧: تكوين المركب الحرفي «إلى المدرسة».

وأن الفعل «حضر» يكون مع الحال «متأخرًا» والمركب الحرفي «إلى المدرسة» المركب الفعلي (VP) كما في الشكل (٤-٨).



الشَّكل ٤-٨: تكوين المركب الفعلي «حضر متأخر إلى المدرسة».

وأخيرًا تتكون الجملة (S) كما في الشكل (٤-٩) من تتابع المركب الاسمي «الطالب» (NP) والمركب الفعلي (VP) «حضر متأخرًا إلى المدرسة».



الشَّكل ٤-٩: الهيكل التركيبي النهائي للجملة في (٧).

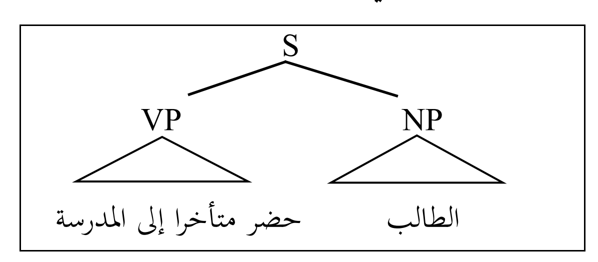
وبشكل عام، يستخدم اللغويون التمثيل الشجري للتعبير عن الهيكل التركيبي للجملة بشكل مفصل. حيث يتم التفريق بين نوعين من الوحدات داخل التمثيل الشجري: النوع الأول يسمى بالوحدات النهائية وهي الكلمات في نهاية الشجرة التحليلية التي لا يمكن تحليلها إلى وحدات أقل على مستوى التركيب مثل (حضر - إلى مدرسة ... إلخ) أما النوع الثاني الوحدات غير النهائية وتشمل كل الوحدات داخل التركيب التي لم يصل فيها التحليل إلى الكلمة المفردة الموجودة داخل المعجم مثل (S – P – V – N – NP – VP ......).

من الأنظمة الأخرى المتبعة في تمثيل التحليل التركيبي للجمل في اللغات الطبيعية استخدام الأقواس للتعبير عن هيكل الجملة. فالجملة في (٧) يمكن تمثيلها كما في الشكل (٤-١٠):

S[VP[PP[NP[N[مدرسة] DET[ال] P[إلى] ADV[متأخرا] V[حضر] NP[N[طالب] DET[ال]]

الشَّكل ٤-١٠: تمثيل التحليل التركيبي للجملة (٧) باستخدام الأقواس.

الشكلان (٤-٩) و(٤-١٠) يوضحان التمثيل التفصيلي للتركيب النحوي للجملة (٧) إلا أن اللغويون يفضلون تمثيل التركيب النحوي للجمل بشكل غير تفصيلي وهو ما يعرف باسم التمثيل الجزئي للتركيب النحوي والذي يتم فيه تمثيل الوحدات الرئيسية التي تُشكل الجملة دون النظر إلى التركيب الداخلي لتلك الوحدات كما في الشكل (٤-١١) والذي استخدم المثلث لتمثيل الوحدات التي تتكون من تركيب معقد والتي لم يتم تمثيلها، وفي الجملة (٨) اتبعت طريقة «استخدام الأقواس» لكن دون تفاصيل وسوف نتناول بشكل مفصل التمثيل الجزئي في الجزء الخاص بـ «دقة المحللات».



الشَّكل ٤-١١: التمثيل الجزئي للتركيب النحوي للجملة (٧).

[[الطالب]NP [حضر متأخرا إلى المدرسة] S VP]   (٨)

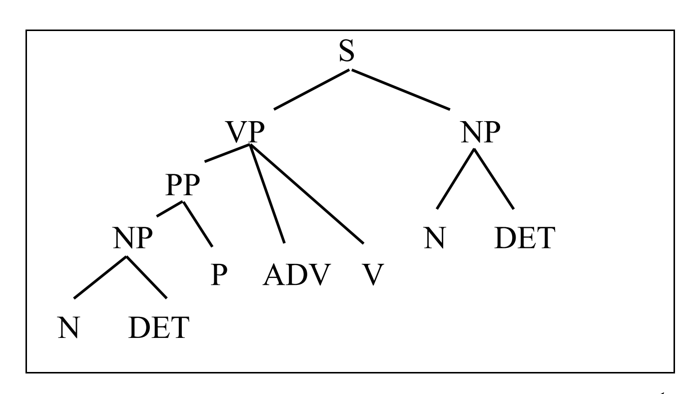
كما ذكرنا من قبل أن القواعد النحوية يجب أن تكون قادرة على توليد وتحليل الجمل الصحيحة، وتعيين البنية النحوية المناسبة للجمل المعنية بالتحليل وكما تبين لنا إنه يمكن تمثيل البنية النحوية للجمل عن طريق التمثيل الشجري كما في الشكل (٤-٩)، فعلينا أن نتساءل عن ماهية القواعد التي تستطيع تمثيل البنية النحوية للجمل، وللجواب عن هذا بالنظر إلى التمثيل الشجري في الشكل (٤-٩) يمكن صياغة مجموعة من القواعد كما في (٩).

(٩) (أ) S → NP VP   (ب) VP → V ADV PP

(ج) PP → P NP   (د) NP → DET N

فالقاعدة (أ) تفترض إمكانية تكوين الجملة من مركب اسمي (NP) متبوع بمركب فعلي (VP). والقاعدة (ب) تفترض أن المركب الفعلي (VP) يتكون من فعل (V) وحال (ADV) بالإضافة إلى مركب حرفي (PP). والقاعدة (ج) تفترض أن المركب الحرفي (PP) يتكون من حرف (P) متبوع بمركب اسمي (NP). ثم أخيرًا القاعدة (د) تفترض أن المركب الاسمي (NP) يتكون من أداة تعريف (DET) واسم (N). وقد أطلق تشومسكي على تلك القواعد قواعد التركيب العباري ((Phrase structure Grammar (PSG) وسميت بذلك لأنها تستطيع تحديد كيفية تكوين الجمل من مركبات وكيفية تكوين المركبات من كلمات، وهذه الافتراضات تلقي الضوء على دور تشومسكي في تطوير علم اللغة الرياضي وتطوير قواعد صورية يمكن من خلالها صياغة قواعد النحو في صورة قواعد رياضية.

وبتطبيق تلك القواعد التي تعتمد على الفئات النحوية (Categorial grammar) في (٩) وحدها نستطيع الحصول على التمثيل الشجري في الشكل (٤-١٢):



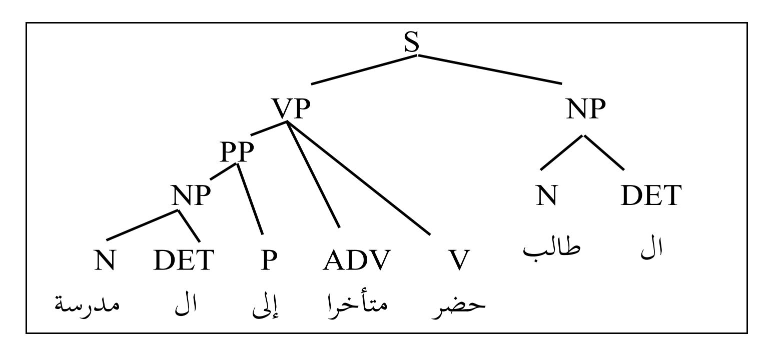
الشَّكل ٤-١٢: التمثيل الشجري للجملة (٧) باستخدام القواعد في (٩).

بالإضافة إلى تلك القواعد التي تعتمد على الفئات النحوية والتي استطاعت تكوين التمثيل الشجري في الشكل (٤-١٢) فالمكون النحوي يحتوي أيضًا على معجم به قائمة من مفردات اللغة مع معلومات عن فئة تلك المفردات ويمكننا هنا أن نتخيل أن المعجم المستخدم لدينا في جملة (٧) يحتوي على المفردات المبينة في الشكل (٤-١٣).

حضر	V
طالب	N
مدرسة	N
متأخرا	ADV
إلى	P
ال	DET

الشَّكل ٤-١٣: معجم به قائمة من مفردات اللغة وفئتها.

وما علينا الآن إلا إدراج المفردات المتواجدة في المعجم والتي لها فئة محددة تحت الفئة التي تنتمي إليها داخل التمثيل الشجري في الشكل (٤-١٢) والمسئول عن ذلك قاعدة تُعرف باسم قاعدة إدراج المفردات وبذلك نتمكن من تمثيل الجملة كما في الشكل (٤-١٤):



الشَّكل ٤-١٤: التمثيل الشجري للجملة (٧) باستخدام القواعد في (٩) مع إدراج المفردات.

وفي ختام هذا العرض ننتهي إلى أن المكون النحوي للغة يحتوي على قواعد تعتمد على الفئات النحوية قادرة على تمثيل التركيب الداخلي للجمل والمعجم وقاعدة إدراج المفردات وتلك المكونات قادرة على توليد التمثيل الشجري في الشكل (٤-١٤) للجملة في (٧) وهو نفسه الموجود في الشكلان (٤-٩) (٤-١٠). ومن هنا نود أن ننوه إننا قد حققنا الهدف المنشود من بناء تلك القواعد النحوية في (٩) والتي استطاعت تحديد البنية التركيبية للجملة (٧)كما هو موضح في الشكل (٤-١٤).

١,٣- القدرة التوليدية للقواعد وقياسها

تتميز اللغات الطبيعية بالقدرة الإبداعية العالية، فالقواعد التي تم بناؤها في (٩) سوف تكون قادرة على توليد عدد لا نهائي من الجمل وليس فقط الجملة الممثلة في الشكل (٤-٩) فكلما زادت عدد القواعد وأيضا زاد حجم المفردات المدرجة داخل القاموس تزداد عدد الجمل التي يمكن توليدها؛ فعلى سبيل المثال إذا تم دخول كلمات - مثل فصل، ولد، صبي، ذهب - وصل.. إلخ - مع وجود نفس القواعد في (٩) فإنه يمكن توليد جمل في الشكل (٤-١٥).

الطالب حضر متأخرا إلى المدرسة.	الولد حضر متأخرا إلى المدرسة.
الصبي حضر متأخرا إلى المدرسة.	الطالب حضر متأخرا إلى الفصل.
الولد حضر متأخرا إلى الفصل.	الصبي حضر متأخرا إلى الفصل.
الطالب ذهب متأخرا إلى المدرسة.	الولد ذهب متأخرا إلى المدرسة.
الصبي ذهب متأخرا إلى المدرسة.	الولد وصل متأخرا إلى الفصل.
الطالب وصل متأخرا إلى الفصل.	الصبي وصل متأخرا إلى الفصل.

الشَّكل ٤-١٥: الجمل التي يمكن توليدها من القواعد في (٩).

وإذا تم تعديل القواعد في (٩) إلى (١٠) بحيث يكون وجود الحال في المركب الفعلي غير إلزامي - وذلك عن طريق إضافة قوسين كما نرى في (ب١٠) مع وجود نفس المعجم الذي استطاع توليد الجمل في الشكل (٤-١٥) فإنه يمكن توليد الجمل كما في الشكل (٤-١٦).

(١٠) (أ) S → NP VP (ب) VP → V (ADV) PP

(ج) PP → P NP (د) NP → DET N

الطالب حضر متأخرا إلى المدرسة.	الولد حضر إلى الفصل.
الصبي وصل متأخرا إلى المدرسة.	الصبي حضر إلى المدرسة.
الولد ذهب متأخرا إلى الفصل.	الطالب حضر إلى المدرسة.
الصبي حضر متأخرا إلى الفصل.	الولد حضر إلى المدرسة.
الطالب حضر متأخرا إلى الفصل.	الطالب حضر إلى الفصل.
الولد حضر متأخرا إلى المدرسة.	الصبي حضر إلى الفصل.

الشَّكل ٤-١٦: الجمل التي يمكن توليدها من القواعد في (١٠).

لذا نحن بحاجة إلى اتساع في المعجم واتساع في عدد القواعد التي تعتمد على الفئات النحوية لتوليد عدد لا نهائي من جمل اللغة الصحيحة، ولا بد من قياس كفاءة القواعد التي تم بناؤها وذلك عن طريق توصيف معايير محددة للتأكد من كفاءة القواعد وقدرتها على وصف تراكيب الجمل وصفًا صحيحًا من أجل محاكاة أفضل للنموذج اللغوي البشري. فقام تشومسكي بوضع ثلاثة معايير أساسية لقياس كفاءة القواعد النحوية لأي لغة وهي: معيار كفاءة الرصد (Observational Adequacy)؛ توصف قواعد اللغة بأنها قد حققت معيار كفاءة الرصد إذا كانت هذه القواعد قادرة على التنبؤ بصحة أو خطأ الجمل نحويًا ودلاليًا وفونولوجيًا وهو أقل المعايير كفاءة. ومعيار كفاءة الوصف (Descriptive Adequacy)؛ تُحقق القواعد معيار كفاءة الوصف إذا كانت قادرة على التنبؤ بصحة أو خطأ الجمل نحويًا ودلاليًا وفونولوجيًا (كفاءة الرصد)، بالإضافة إلى قدرتها على أن تصف بشكل صحيح نحويًا ودلاليًا وفونولوجيًا

بنية الجمل في اللغـــة وتــقدم وصفًا يعطــي تفسيرًا وصفيًا لتراكيـــب اللغــة كما في النموذج البشري لابن اللغة. ومعيار كــفاءة التفسير (Explanatory Adequacy) هو أعلى معيار يحدد كفاءة القواعد. ويتحقق معيار كفاءة التفسير إذا كانت القواعد قادرة أن تحقق درجـــة كفاءة الوصف بالإضافة إلى شرطين أساسين. الأول: إذا كانت القواعد لا تحتوي على أي قاعدة تتنافى مع مبدأ كتابة القواعد في أي لغـــة من اللغـــات أي أن القواعد تستطيع وصـــف جمـــل اللغـــة في ظـــل وجود مجمـــوعة من الشـــروط المحـــددة والمتفـــق عليهـــا عالميـــا في كتـــابة القواعد.

فعلى سبيل المثال فإنه من غير المتوقع أن تحتوي أي لغة على قاعدة نحوية تنص على أنه يمكن أن تحلل الجملة إلى مركب اسمي ومركب فعلي في السادسة مساءً؛ فتلك القاعدة لا يمكن أن تتواجد في أي لغة من اللغات الطبيعية وذلك لأنها تعتمد على معلومات غير لغوية (الزمن) وذلك منافٍ لشروط بناء القواعد في أي لغة من اللغات (Chomsky ١٩٥٦). أما الشرط الثاني فهو أن تحاكي القواعد الطبيعة النفسية الواقعية والقدرة الذهنية للنموذج البشري للغة.

٢,٣- هل تبنى قواعد التركيب على فكرة وجود المركبات والفئات النحوية أم على فكرة وجود الكلمات؟

تتسم القواعد التي تعتمد على الفئات النحوية بقدرتها على تحقيق مبدأ التعميم. فمنظومة النحو تهتم ببناء نظام قادر على تحديد توزيع وترتيب الكلمات داخل تركيب الجمل. فأي محاولة لبناء نظام لا يعتمد على الفئات النحوية لن يكون قادرًا على بناء قواعد لتكوين المركبات وتوزيعها داخل الجمل واستخدام هذه القواعد لتوليد عدد لا نهائي من الجمل. فإذا رفضنا فكرة القواعد المبنية على المركبات التي تعتمد على الفئات النحوية، فلن يكون هناك وسيلة لبناء قواعد لتحليل وتوليد مجموعة الجمل مثل تلك الموجودة في (١١) إلا باستخدام قواعد تبنى على الكلمة ، وهذا ما يعرف بـ (Noncategorial Grammar) أو (Word-based Grammar).

المدرس الماهر قدم طالب ناجح للمدير الفاضل. (١١)

الطالب الماهر قدم مدير فاضل للمدرس الناجح.

المدير الناجح وصف مدرس ماهر للطالب الذكي.

الطالب الذكي قدم مدير ناجح للمدرس الفاضل.

الطالب الماهر قدم مدرس ذكي للمدير الناجح.

وفي هذه الحالة يتم توليد الجمل في (١١) عن طريق عمل مصفوفة تضم مكان لكل كلمة من كلمات الجملة وتوضع كل كلمة في عمود كما نرى في الشكل (٤-١٧):

١	٢	٣	٤	٥	٦	٧	٨	٩	١٠	١١	١٢
ال	مدرس	ال	ماهر	قدم	طالب	ناجح	ل	ال	مدير	ال	فاضل
	طالب		.....	وصف	مدير	فاضل			طالب		ناجح
	مدير		ناجح	.....	مدرس	ماهر			مدرس		ذكي
	....		ذكي	.....	.....	.....			.....		فاضل
	.....		فاضل	.....	.....	ذكي			.....		......
	.....		.....	.....	.....	.....			.....		.....

الشَّكل ٤-١٧: مصفوفة تضم مكان لكل كلمة من كلمات الجملة.

وعندها يكون بناء مجموعة من القواعد لتوليد أي جملة عن طريق اختيار كلمة من عمود رقم (١) ثم من عمود رقم (٢) ثم من عمود (٣)...... وهكذا. وهذا الاتجاه في بناء القواعد يؤدي إلى مشكلات منهجية عدة منها: (أ) التخلي عن مبدأ التعميم في القواعد؛ فلا بد من ذكر كل كلمة بعينها بدلا من ذكر فئتها. (ب) نجد تكرارا واضحا بين الكلمات وارتباطها بمكان توزيعها؛ فمجموعة الكلمات في العمود رقم (٢) هي نفسها الكلمات في العمود رقم (٦) ورقم (١٠). (ج) لا يمكن لهذه المنهجية التعامل مع ظاهرة التكرار في بناء المركبات داخل الجملة؛ يمكن للجملة أن يزداد طولها إلى ما لا نهاية بسبب العطف مثلا كما في (١٢):

رأيت محمد وعلي وأحمد وخالد وطارق وعمر. (١٢)

فإن أي محاولة لبناء قواعد لا تعتمد على الفئات النحوية لوصف ظاهرة التكرار في اللغة بشكلها المطلق ستكون مستحيلة. وإلا سنضطر إلى بناء مصفوفة من أعمدة لا نهائية من الكلمات كلما تكرر العطف. فيستحيل على هذا النوع من القواعد تفسير القدرة التوليدية اللا نهائية للجمل. وسيؤدي ذلك بدوره إلى خرق المبدأ الرئيسي لبناء القواعد الذي لا بد فيه أن تتوافر إمكانية بناء عدد محدود من القواعد قادرة على توليد عدد لا نهائي من الجمل، ولكن في هذه الحالة ستكون عدد القواعد لا نهائية مع زيادة عدد الأعمدة في المصفوفة الذي يعتبر أمرا غير مقبول كمبدأ بناء للقواعد اللغوية.

ولكن إذا ما اعتمدنا في نظام القواعد على الفئات النحوية فنجد أن جميع المشكلات السابقة لن يكون لها وجود. فظاهرة التكرار يمكن التعامل معها بسهولة في نظام القواعد التي تعتمد على الفئات النحوية وتفسير إمكانية توليد عدد لا نهائي من الكلمات طبقا للقواعد الموصفة. ومن نقاط القوة أيضًا التي ترجح كفة نظام القواعد التي تعتمد على الفئات النحوية أن هذه القواعد تمثل نظامًا أكثر إحكامًا وتقييدًا من نظام القواعد التي لا تعتمد على الفئات النحوية. فنظام القواعد التي تعتمد على الفئات النحوية يفترض وجود قيود تحدد القدرة التوليدية للقواعد شرط أن تكون مكونة من عدد محدد من الفئات العبارية (Phrasal Categories) والفئات المعجمية (Lexical Categories) وهي (N, NP, V, VP, ADV, ADVP ...) ولا يمكن الخروج عن هذه الفئات (لمزيد من المعلومات يمكن الاطلاع على (Radford, 1988)).

إذًا نستنتج مما سبق أن نظام القواعد التي تعتمد على الفئات النحوية هو الأكثر تعبيرًا عن طبيعة بناء القواعد وذلك لثلاثة أسباب: أولاً قدرته على التعميم وثانيًا قدرته على التعامل مع القدرة التوليدية اللا نهائية للجمل وأخيرًا طبيعته الأكثر إحكامًا وتقييدًا. فلا بد لأي عمل يسعى إلى بناء قواعد نحوية قادرة على توليد وتحليل الجمل العربية أن تستمر في هذه المهمة لنصل إلى قواعد نحوية عالية القدرة الإنتاجية في توليد وتحليل عدد لا نهائي من الجمل.

٣,٣- البنية السطحية والبنية العميقة للجملة

إن قواعد النحو لا بد أن تكون قادرة على توليد كل البنى السطحية للجمل في اللغة وهذا يتحقق عن طريق مجموعة من القواعد التي تعتمد على الفئات النحوية والمعجم وقاعدة إدراج المفردات. ولكن تلك المكونات وحدها لم يكن لديها القدرة على التعامل مع أنواع أخرى من التراكيب تظهر في اللغات الطبيعية؛ فعلى سبيل المثال الجملة المبنية للمجهول في (١٣):

الأقلام وُضعت في الحقيبة (١٣)

فإن القواعد التي تعتمد على فئة المفردة التي تم بناؤها سوف تتنبأ خطًا أن الجملة في (١٣) غير صحيحة، وذلك لأنها تنتهك الإطار النحوي (Subcategorisation frame)() المبني للفعل «وضع»؛ فالفعل «وضع» يحتاج إلى متعلقات نحوية لإتمام تركيب الجملة بشكل صحيح، كقولنا «الرجل وضع الأقلام في الحقيبة» فيأتي ورائه مركب اسمي ومركب حرفي ولكن في الجملة (١٣) نجد أن الفعل (وضع) يأتي ورائه مركب حرفي فقط ولا يأتي ورائه مركب اسمي.

ولكن الجملة في (١٣) هي جملة صحيحة ولكن حدث فيها تحريك للعبارة الاسمية (الأقلام) عن طريق عملية تعرف بعملية تحريك المركب الاسمي؛ حيث إنه يتم تحريك المركب الاسمي من مكانه الأصلي (بعد الفعل) في البنية العميقة للجملة كما في (١٤) وهنا يجب أن نفرق بين البنية العميقة للجملة وهي الشكل الباطني وغير الظاهر للجملة والبنية السطحية وهي على عكس البنية العميقة.

________ وضع الأقلام في الحقيبة. (١٤)

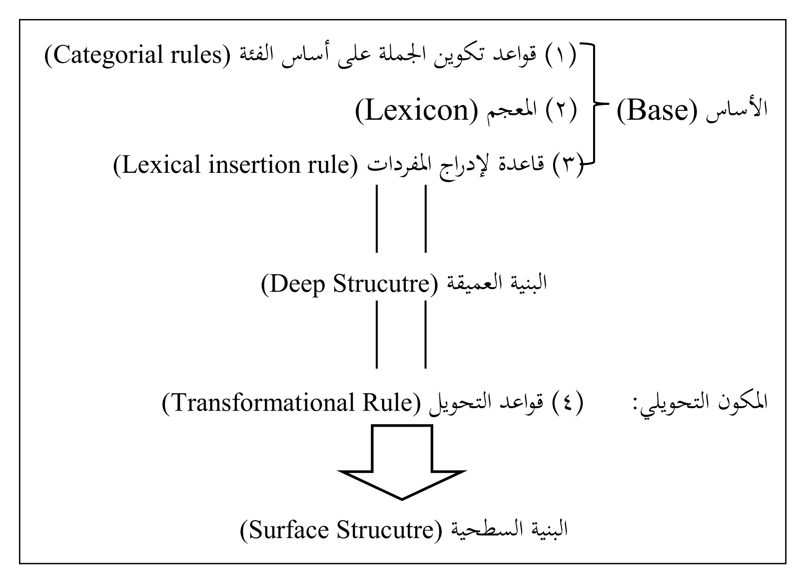
في (١٤) استخدام (____) يعبر عن موضع الفاعل ونجد أن الفعل (وضع) يقع داخل المركب الفعلي متبوعا بالمركب الاسمي (الأقلام) متبوعا بالمركب الحرفي (في الحقيبة)، ومن هنا يمكن لنا أن نستنتج أن الجملة في (١٣) نشئت من البنية الأساس (Base) في (١٤) التي تم توليدها خلال المكون النحوي المبني على أساس القواعد والمعجم وقاعدة إدراج المفردات فقط.

وبناء القواعد بهذه الطريقة يعاني من مشكلات أخرى، فعلى سبيل المثال سوف يتمكن المكون النحوي من توليد الجملة في (١٥) ولكنه لا يتمكن من توليد الجملة في (١٦).

الرجل أكل التفاحة. (١٥)

التفاحة أكلها الرجل. (١٦)

وذلك لأن تلك القواعد تستطيع أن تقدم الوصف الذي يظهر فيه المتعلقات النحوية لكل مسند في أماكنها المناسبة. ومن هنا كان لا بد من نشأة نوع آخر من القواعد هي قواعد التحويل؛ حيث تتحول الجملة (١٦) من الجملة (١٥) بواسطة إجراء تحويل ينقل المركب الاسمي إلى موقع الابتداء وهذا التحويل يتناول نقل الركن الاسمي الذي يقع على يسار الفعل إلى موقع الابتداء تاركا مكانه عائدا يعود عليه في الموقع الذي كان يشغله (زكريا، ١٩٨٣). ولذلك لا بد من إعادة تنظيم النحو التحليلي لإدراج مكون تحويلي كما في الشكل (٤-١٨):



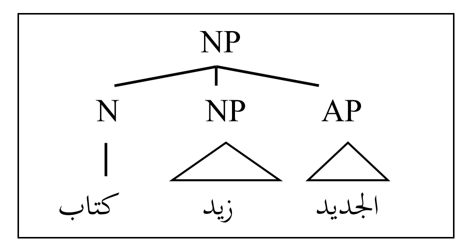
الشَّكل ٤-١٨: منظومة النحو التحليلي بناء على وجود مكون تحويلي.

وهكذا يمكن لقواعد النحو في وضعها الحالي أن تكون قادرة على توليد جمل من اللغة وهذا يتحقق عن طريق مجموعة من القواعد والمعجم وقاعدة لإدراج المفردات

وأخيرا قواعد التحويل. وعلينا أن نستمر في تحسين قدرة النحو من أجل الوصول إلى قواعد على درجة عالية من الكفاءة قادرة عل التعامل مع جميع تراكيب اللغة.

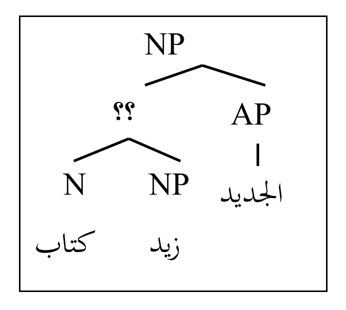
٤,٣- المركبات الوسيطة

إذا نظرنا إلى التتابع «كتاب زيد الجديد» في الشجرة النحوية في الشكل (٤-١٩) سنجد أنه ينقسم إلى ثلاث وحدات (N) و(NP) و(AP).



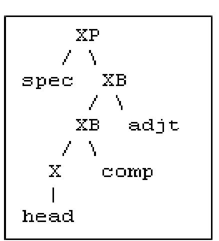
الشَّكل ٤-١٩: الشجرة النحوية لجملة «كتاب زيد الجديد».

ففي هذا التمثيل التركيبي يمكن لنا القول بأن المركب الوصفي «الجديد» في الشكل (٤-١٩) يصف الاسم «كتاب» حيث يخرج فرع المركب الوصفي والاسم «كتاب» من نفس النقطة. ولعل ذلك ليس بالدقة الكافية، لأن المركب الوصفي «الجديد» يصف (كتاب زيد) وليس «كتاب» فقط، كما أن «كتاب زيد» يمثل وحدة أيضا بدليل صحة العطف عليه بمركب من نفس النوع كما في «كتاب زيد وقلم عمرو جديدان»، كما يصح معه الإسناد كما في «هذا كتاب زيد»؛ وذلك لا يظهر في الشكل (٤-١٩). ولذلك لا بد من تعديل التمثيل التركيبي لـ الشكل (٤-١٩) إلى التمثيل الموضح في الشكل (٤-٢٠):



الشَّكل ٤-٢٠: التمثيل التركيبي المعدل لــ الشكل (٥-١٩).

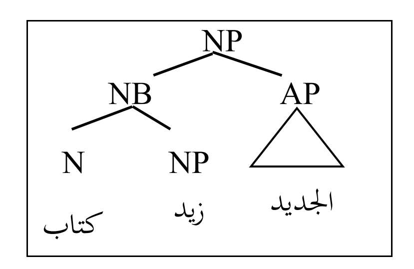
فنجد أن في الشكل (٤-٢٠) المركب الوصفي «الجديد» يصف المركب الجديد «غير المسمى» «كتاب زيد»، وكما ذكرنا أن هذا المركب يمثل وحدة نحوية فلعلنا نتساءل عن فئة هذا المركب «؟؟» والتي تضم التتابع (N – NP) ولعلنا في الوهلة الأولى نظن أنه مركب اسمي (NP) لأن الرأس اسم «كتاب» (N). ولكن لا يمكن أن يكون هذا منطقيا إذ إن هذا يعتبر نقصًا في التوصيف النحوي لأن المركب الاسمي الأعلى لن يكون له رأس. ومن هنا يمكن استنتاج أن هذه العبارة تمثل مركب وسيط، هذا المركب أكبر من الاسم (N) وأصغر من المركب الاسمي (NP). وهنا نجد فجوة كبيرة داخل نظام قواعد التركيب العباري (PSG) للتعبير عن هذا المركب فلا توجد فئات وسيطة ما بين الفئات المعجمية والفئات العبارية في إطار عمل تلك النظرية في الوصف التركيبي للجملة مما يؤدي إلى عدم قدرة أي نحو توليدي من الوصول إلى مستوى «كفاءة الوصف» حيث أنه لن يكون قادرا على وصف هذا النوع من التراكيب في «كتاب زيد الجديد». فظهر مفهوم نظرية X–bar؛ ووُضعت هذه النظرية عام ١٩٧٠ بواسطة نعوم تشومسكي لتعالج النقص الموجود في التوصيف النحوي في PSG وتم تطويرها فيما بعد بواسطة راي جاكندوف عام ١٩٧٧. وتنص هذه النظرية على أن كل اللغات الطبيعية البشرية تتبادل تراكيب متشابهة محددة لها نفس البنية العميقة والتي يمكن تمثيلها في الشكل (٤-٢١).



الشَّكل ٤-٢١: تمثيل البنية العميقة باستخدام نظرية X-bar.

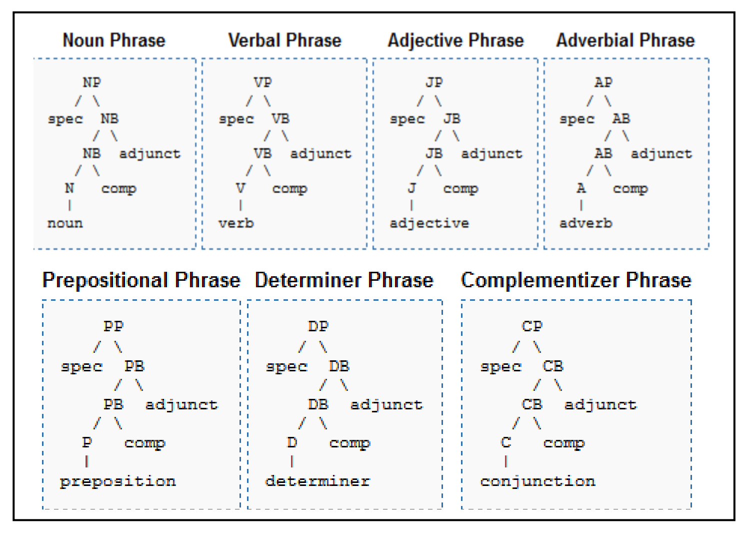
حيث يُستخدم الحرف (X) في هذه النظرية لكي يعرف الفئات المعجمية الأساسية، أو رؤوس العبارات؛ فهناك أنواع من رؤوس العبارات وهي (N) للأسماء مثل الضمائر الشخصية وأسماء الإشارة .... إلخ و(V) للأفعال و(J) للصفات و(A) للظروف و(P) لحروف الجر و(D) للمحددات كالأدوات على سبيل المثال و(C) لحروف العطف. وتحدد رؤوس العبارات نوع أو فئة العبارات فنجد أن N تنتج مركب اسمي

((NP، V تنتج مركب فعلي (VP) وJ تنتج مركب وصفي ((Adjective Phrase JP)) وA تنتج مركب ظرفي ((Adverb Phrase (AP) وP تنتج مركب حرفي (PP) وD تنتج عبارة تحديدية ((Determiner Phrase(DP) وC تنتج مركب تكميلي ((Complementizer Phrase(CP). وفي شكل (٤-٢١) الرمزين XP وXB يسميان إسقاطين (Projections)، والرمز X يسمى رأس العبارة، ويسمى أيضًا بالإسقاط الصفري (Zero Projection) حيث أنه الوحدة النهائية ولا يمكن إسقاط أي وحدات نحوية منه وعلى هذا فإن كل عبارة لها إسقاط صفري واحد أي رأس واحدة. والوحدة العليا المسماة بـ XP تسمى بالإسقاط الأقصى (Maximal Projection) حيث أنه أكبر وحدة نحوية يمكن أن يُدرج تحتها أي عدد من الإسقاطات. فكل الإسقاطات الواقعة بين الرأس والإسقاط الأقصى تسمى بالإسقاطات الوسيطة ويمكن وصف الإسقاطات الوسيطة بالوحدات غير النهائية وذلك لإمكانية إسقاط وحدات نحوية منها؛ حيث أن الإسقاط الأقصىXP يسيطر على الإسقاط الوسيط XB الذي يسيطر على الإسقاط الصفري X. أما «Specifier) «spec) أو «المخصص النحوي» وهي الكلمة أو العبارة أو الجملة التي تحدد العنصر «X» أما « Complement)) «comأو «المكمل» وهو العنصر النحوي اللازم لإكمال سلامة التركيب ، أما «Adjunct) «adjt) أو «الملحق» فهي الكلمة أو العبارة أو الجملة التي تصف العنصر «X» ولكن ليس من الضروري وجوده نحويا. ويمكننا رؤية التمثيل النحوي لأشكال المركبات المختلفة في الشكل (٤-٢٢).



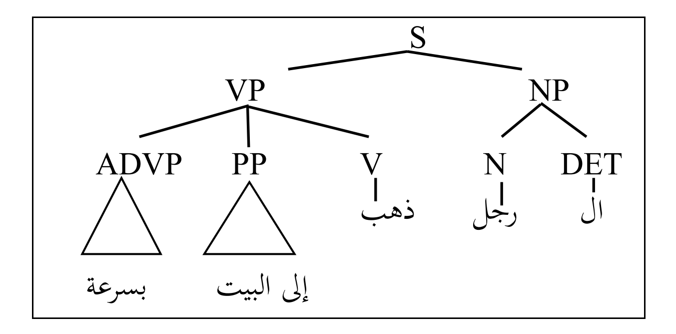
الشَّكل ٤-٢٢: التمثيل النحوي لأشكال المركبات باستخدام نظرية X-bar.

وبعد أن تعرفنا على فكرة وجود المركب الوسيط من خلال نظرية X bar يمكننا إعادة تمثيل في الشكل (٤-٢٠) إلى الشكل (٤-٢٣):



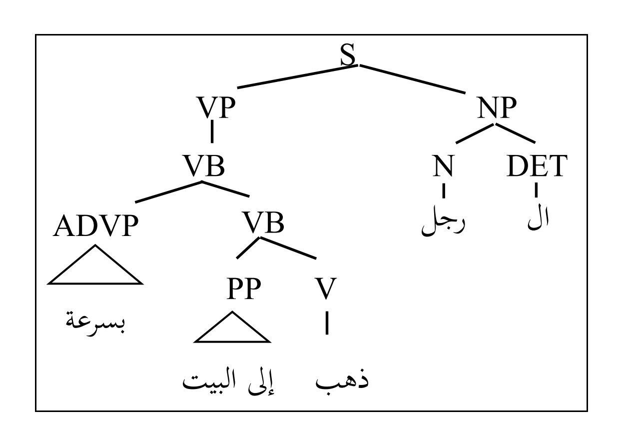
الشَّكل ٤-٢٣: تعديل التمثيل التركيبي لــلشكل (٥-٢٠) باستخدام المركب الوسيط.

لمزيد من ترسيخ مفهوم المركبات الوسيطة يمكننا أخذ مثال آخر فإذا نظرنا إلى الجملة «الرجل ذهب إلى البيت بسرعة»، نجد أن تمثيل الجملة على النحو الموضح في الشكل (٤-٢٤) لا يعبر عن حقيقة أن المركب (بسرعة) يصف الحدث (ذهب إلى البيت).



الشَّكل ٤-٢٤: التمثيل الشجري لجملة «الرجل ذهب إلى البيت بسرعة» باستخدام قواعد التركيب العباري (PSG).

وهنا لا بد أن نجمع بين الفعل (ذهب) والمركب الحرفي (إلى البيت) اللذان يعبران معا عن وحدة واحدة - حيث أن المركب (إلى البيت) هو المكمل للفعل (ذهب) - لذلك يأتي دور المركبات الوسيطة للتعبير عن ذلك كما في الشكل (٤-٢٥):



الشَّكل ٤-٢٥: ربط الفعل (ذهب) والمركب الحرفي (إلى البيت) في وحدة واحدة (VB).

حيث نجد أن المركب الفعلي يتكون من إسقاط وسيط والذي يتكون بدوره من إسقاط وسيط آخر (يتكون من الفعل والمركب الحرفي) ومركب ظرفي. أي أن المركب الظرفي أصبح يصف المركب (ذهب إلى البيت).

وبذلك يمكننا القول بأن قواعد النحو أصبحت الآن على درجة أعلى من الكفاءة ويمكنها الوصول إلى مستوى كفاءة الوصف؛ حيث تستطيع تمثيل مستوى أعم من تراكيب اللغة المختلفة وعلى ذلك فإنه ينبغي الاستمرار في تحسين قدرتها أكثر فأكثر حتى تصل إلى درجة عالية من الكفاءة تمكنها من التعامل مع التراكيب اللغوية بشكل كامل.

٥,٣-الفصل بين التركيب والدلالة

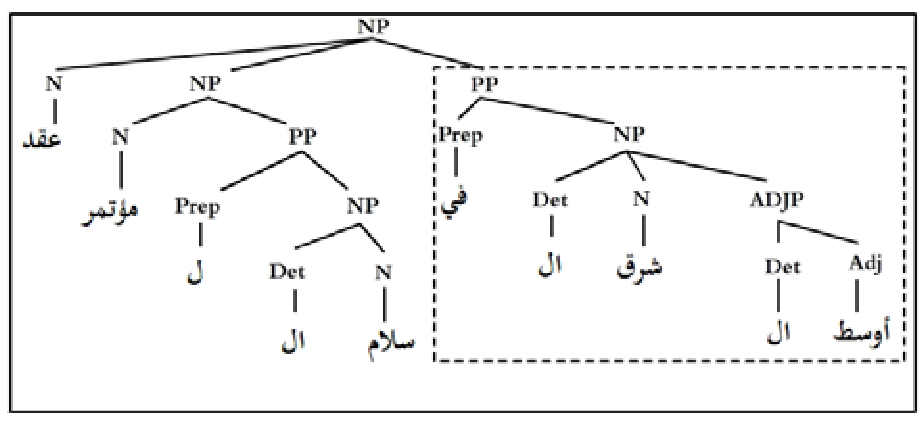
قبل طرح قضية الفصل بين التركيب والدلالة لا بد أن نشير إلى ظاهرة لغوية تتميز بها اللغات الطبيعية وهي الالتباس التركيبي؛ ويقصد هنا باللبس التركيبي أن يكون هناك أكثر من تحليل للجملة بعد تطبيق قواعد التحليل، فعلى سبيل المثال الجملة في (١٧).

عقد مؤتمر للسلام في الشرق الأوسط. (١٧)

فإذا نظرنا إلى القواعد التي سوف يتم تطبيقها للحصول على البنية التركيبية للجملة في (١٧) فنجد أن تطبيق القواعد في (١٨) يصل بنا إلى تمثيل الجملة خلال شجرتين.

(أ) NP→ N NP PP
(ب) NP→ N PP
(ت) PP→ Prep NP
(ث) NP→ Det N (١٨)
(ج) NP→ Det N ADJP
(ح) ADJP→ Det Adj
(خ) NP→ N NP

فعند تطبيق القواعد في (١٨) لتحليل الجملة في (١٧) فإنه يمكن الحصول على البنية الموضحة في الشجرة في الشكل (٤-٢٦).

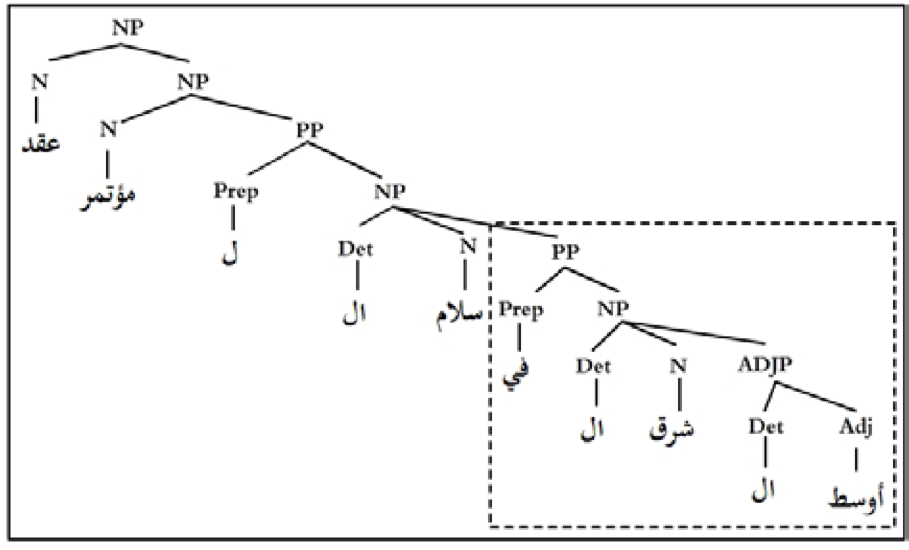


الشَّكل ٥-٢٦: التحليل الشجري للجملة (١٧).

حيث يتركب المركب الاسمي «عقد مؤتمر للسلام في الشرق الأوسط» من اسم (N) «عقد» ومركب اسمي (NP) آخر «مؤتمر للسلام» ومركب حرفي (PP) «في الشرق الأوسط» كما في (١٨-أ) ويتكون المركب الاسمي «مؤتمر للسلام» من اسم «عقد» ومركب حرفي «للسلام» كما في (١٨-ب) ويتكون المركب الحرفي «للسلام» من حرف «ل» ومركب اسمي «السلام» كما في (١٨-ت) ويتكون ذلك المركب الاسمي من أداة تعريف «ال» واسم «سلام» كما في (١٨-ث)، أما المركب الحرفي «في الشرق الأوسط» الممثل داخل المربع في الشكل (٤-٢٦) فهو يتكون من حرف «في» ومركب اسمي

«الشرق الأوسط» وذلك من خلال القاعدة (١٨-ت) والمركب الاسمي «الشرق الأوسط» يتكون من أداة تعريف «ال» واسم «شرق» ومركب وصفي وذلك من خلال القاعدة في (١٨-ج) وأخيرا يتكون هذا المركب الوصفي من أداة تعريف «ال» وصفة «أوسط» من خلال القاعدة في (١٨-ح).

ولكن إذا نظرنا إلى القواعد في (١٨) فإننا نجد إنه يمكن تطبيق القواعد بشكل آخر وعند ذلك يمكننا الحصول على تمثيل مختلف للبنية النحوية للجملة في (١٧) وهو الموضح في الشكل (٤-٢٧) والذي يختلف عن التمثيل في الشكل (٤-٢٦).



الشَّكل ٥-٢٧: تحليل شجري آخر للجملة (١٧).

حيث يتركب المركب الاسمي «عقد مؤتمر للسلام في الشرق الأوسط» من اسم (N) «عقد» ومركب اسمي (NP) آخر «مؤتمر للسلام في الشرق الأوسط» خلال القاعدة (١٨-خ).

فإذا نظرنا إلى الشكل (٤-٢٦) نجد أن المركب الحرفي «في الشرق الأوسط» وصف على أنه يرتبط بـ «عقد المؤتمر»، أي أن عقد المؤتمر تم في الشرق الأوسط، وتم تمثيل ذلك في الشجرة النحوية عن طريق إضافة فرع من المركب الاسمي، وذلك على خلاف ما نجده في الشكل (٥-٢٧)؛ حيث نجد أن الجملة تم تحليلها على أساس أن المركب

الحرفي «في الشرق الأوسط» يرتبط بالمركب الاسمي «السلام»، أي أن «السلام» تم تحديده بأنه في الشرق الأوسط، وتم التعبير عن ذلك في توصيف الشجرة النحوية حيث أن المركب الحرفي متصل بفرع من المركب الاسمي «السلام». وهنا نجد أن اختلاف التركيب النحوي كما في الشكلين (٥-٢٦)، (٥-٢٧) أدى إلى اختلاف في المعنى وتفسير الغموض في الجملة في (١٧) وعلينا أن ننوه أن ما حدث من تفسير للغموض الذي هو ظاهرة دلالية لا يعد تداخل بين التركيب والدلالة لأن التداخل معناه استخدام مفاهيم من مستوى آخر (الدلالة) لوصف تركيب الجملة فما حدث هنا يدل على أن النحو بشكل طبيعي يمكنه أن يفسر المعنى دون استخدام مفاهيم من الدلالة، وتعد تلك القدرة حسب ما وصفه تشومسكي أحد المميزات التي تضاف إلى قدرة القواعد النحوية التي لديها القدرة على تكوين بنى نحوية تفسر معاني الجمل الطبيعية.

ويتفق هذا مع إشكالية ما أسماه تشومسكي باستقلالية النحو (Autonomy of Syntax)، حيث يرى تشومسكي «أن التركيب مستقل عن الدلالة، ولا علاقة له بها» (سيرل، ١٩٧٩)، وإنما يتم النظر إلى منظومة النحو دون الرجوع إلى المعنى أو السياق. ولعلنا هنا قد لفتنا النظر إلى علاقة الـتأثير ما بين منظومة النحو والدلالة. فوجود تفاعل بين المنظومتين لا يعني الدمج بينهما. ولمزيد من التفاصيل عن قضايا اللبس التركيبي يمكنك الاطلاع على (Al-Ansary, & El-Kareh, 2004).

٤- مدخل إلى معالجة تركيب الجملة آليا

بعد أن تعرفنا في الأجزاء السابقة على كيفية تكوين البنية التركيبية للجمل وكيفية بناء قواعد تسمح بتحديد البنية التركيبية للجمل، فنكون بذلك قد ألقينا الضوء على بعض المفاهيم الأساسية عند تناول التحليل النحوي للجمل؛ فالتحليل النحوي يعد من المتطلبات الهامة والمحورية في العديد من تطبيقات معالجة اللغات الطبيعية كالترجمة وأنظمة الإجابة على الأسئلة وتطبيقات نظم الحوار، فمثل هذه التطبيقات تتطلب معرفة معاني جمل اللغة الطبيعية. ولإتقان معاني الجمل لا بد من توفر التركيب النحوي للمكونات الجوهرية للجملة. والهدف الأساسي من التحليل النحوي هو تحديد التراكيب النحوية التي تعبر عن مدى التطابق بين جمل اللغة الطبيعية ومعانيها (Miyao, 2006). فالأبحاث التي تتضمن العلاقة بين الجملة ومعناها تعد من الأبحاث الصعبة

والضرورية لفهم اللغة الطبيعية. فما علينا الآن إلا بناء نظام يمكن أن يقوم بتنفيذ مهمة تكوين البنية التركيبية للجمل آليا، وهو ما يعرف باسم المحلل النحوي حيث إن إتاحة المحللات النحوية الآلية هي نقطة انطلاق للمعالجة الذكية للغات الطبيعية.

١ , ٤ – تصميم منطق بناء المحللات النحوية

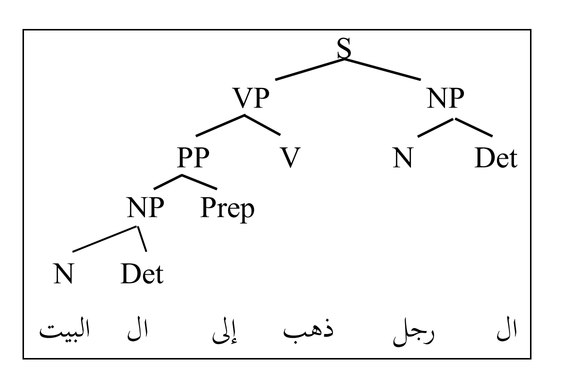
تتعدد الاتجاهات والتقنيات المتبعة عند بناء المحللات النحوية فيتم بنائها بالاعتماد على اتجاهين للتحليل النحوي؛ التحليل النحوي من أعلى إلى أسفل والتحليل النحوي من أسفل إلى أعلى ويمكن من خلال هذين الاتجاهين الوصول إلى نفس التركيب النحوي للجملة على الرغم من الاختلاف في طريقة تطبيق وترتيب القواعد في كلا الاتجاهين، وعلى المحلل النحوي أن يحدد ما إذا كان الاتجاه المتبنى للتحليل غير محدد أو محدد (Dobrovolsky, Katamba, & O'Grady, 1997). وفيما يلي سوف نتناول عرض كل اتجاه بالتفصيل ويلي ذلك توضيح لفكرة التحديد وعدم التحديد في الاتجاه المتبنى أثناء التحليل النحوي.

- **التحليل النحوي من أعلى إلى أسفل**

التحليل النحوي من أعلى إلى أسفل هو أحد طرق بناء المحللات النحوية ويتم فيه بناء الشجرة النحوية من خلال النظر إلى الجملة (أعلى مستوى) وتجزئتها عن طريق القواعد النحوية، يتم خلال هذا النوع من التحليل التعرف على العلاقات بين أجزاء المركب.

وفي هذا الاتجاه يبدأ المحلل النحوي بتطبيق القواعد لبناء شجرة للجملة المدخلة بدءً من الوحدات غير النهائية وانتهاءً بالوحدات النهائية؛ أي مفردات الجملة، ومرورًا بعدد من القواعد والقيود التي تعمل على اختيار الكلمة المناسبة وتسكينها في المكان الملائم لها في هيكل الشجرة النحوية وفقًا لفئة الكلمة. فعلى سبيل المثال إذا أدخلنا جملة «الرجل ذهب إلى البيت» فيتم تطبيق عدد من القواعد كما هو في (١٩) لبناء الشجرة النحوية في الشكل (٤-٢٨) لهذه الجملة وسوف نقوم بعرض تلك القواعد بالتفصيل:

(١٩)
(أ) S
(ب) NP VP
(ت) Det N Verb PP
(ث) Det N Verb Prep NP
(ج) Det N Verb Prep Det N
(ح) ال N Verb Prep Det N
(خ) ال رجل Verb Prep Det N
(د) ال رجل ذهب Prep Det N
(ذ) ال رجل ذهب إلى Det N
(ر) ال رجل ذهب إلى ال N
(ز) ال رجل ذهب إلى ال بيت



الشكل ٤-٢٨: بناء الشجرة النحوية لجملة «الرجل ذهب الى البيت» من أعلى إلى أسفل.

عند دخول الجملة - ويشار لها بـ S - إلى المحلل، فإن المحلل يستوعب ذلك عن طريق القاعدة في (١٩-أ) وبعد ذلك يتم تحديد أن هذه الجملة تتكون من مركب اسمي وآخر فعلي كما في (١٩-ب)، وأن المركب الاسمي الذي تم الحصول عليه خلال (١٩-ب) يتكون من أداة تعريف واسم كما هو مبين في (١٩-ت) وكذلك المركب الفعلي في (١٩-ب) يتكون من فعل ومركب حرفي كما في (١٩-ث) وأن المركب الحرفي يتكون من حرف ومركب اسمي كما هو مبين في (١٩-ج) وهذا المركب الاسمي في (١٩-ج) يتكون من أداة تعريف واسم كما في (١٩-ح)؛ وعند هذه المرحلة يكون قد تم الوصول إلى العلاقة البنائية لأجزاء الجملة؛ وتم أيضا تحديد الفئة النحوية لكل جزء في الجملة. أما في القواعد (١٩-خ)، (١٩-د)، (١٩-ذ)، (١٩-ر)، (١٩-ز)، (١٩-س) فيتم فيه تسكين كل مفردة من مفردات الجملة في المكان المناسب لها داخل الشجرة النحوية التي تم بناء فروعها.

■ **التحليل النحوي من أسفل إلى أعلى**

هذا التحليل هو المقابل للتحليل السابق؛ حيث يقوم المحلل النحوي بتقسيم الجملة المدخلة إلى مفردات ووحدات أصغر هي نقطة الانطلاق في التحليل؛ ففي هذا الاتجاه يبدأ المحلل النحوي بتطبيق القواعد لبناء شجرة للجملة المدخلة بدءًا من الوحدات النهائية وانتهاءً بالوحدات غير النهائية. فيتم تقسيم الجملة إلى كلمات وكل كلمة يتم وسمها بالفئة النحوية التي تنتمي إليها ويتم استخدام هذه الفئات لبناء علاقات تركيبية بين الكلمات بعضها البعض لتكوين مركبات أكبر، يتم الربط بين هذه المركبات لبناء الشجرة التحليلية الكاملة للجملة. فعلى سبيل المثال إذا أدخلنا جملة «الرجل ذهب إلى البيت»؛ فإن المحلل يقوم أولاً بالتعرف على المفردات المكونة للجملة من واقع المعجم وهي «ال» – «رجل» – «ذهب» – «إلى» – «ال» – «بيت» وهي على الترتيب أداة تعريف، اسم، فعل، حرف، أداة تعريف، اسم. ويمكن توصيف مراحل التحليل بشكل مبسط عن طريق عدد من القواعد الموصفة في (٢٠) لاستخراج الشجرة في الشكل (٤-٢٩):

(٢٠)

(أ) ال رجل ذهب إلى ال بيت

(ب) Det رجل ذهب إلى ال بيت

(ت) Det N ذهب إلى ال بيت

(ث) Det N Verb إلى ال بيت

(ج) Det N Verb Prep ال بيت

(ح) Det N Verb Prep Det بيت

(خ) Det N Verb Prep Det N

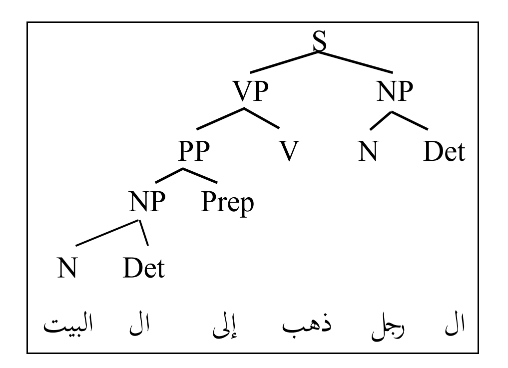
(د) NP Verb Prep Det N

(ذ) NP Verb Prep NP

(ر) NP Verb PP

(ز) NP VP

(س) S



الشكل ٤-٢٩: بناء الشجرة النحوية لجملة «الرجل ذهب الى البيت» من أسفل إلى أعلى.

ففي القاعدة (٢٠-أ) يتم التعرف على المفردات المكونة للجملة وبعد ذلك يتم وسم تلك المفردات كما في القواعد رقم (٢٠-ب)، (٢٠-ت)، (٢٠-ث)، (٢٠-ج)، (٢٠-ح)، (٢٠-خ) بعد ذلك يبدأ المحلل في تكوين مركبات أكبر ففي القاعدة (٢٠-د) و(٢٠-ذ) يتم تكوين مركب اسمي من خلال الدمج بين أداة التعريف والاسم الذي تم الحصول عليهم في (٢٠-خ). وكذلك من خلال (٢٠-ر) يتم تكوين المركب الحرفي عن طريق الجمع بين حرف الجر والمركب الاسمي الذي يليه والذي قد تم الحصول عليهم من خلال القاعدة (٢٠-ذ). وبعد ذلك يتم الدمج بين المركب الحرفي والفعل اللذين قد تم الحصول عليهم خلال القاعدة (٢٠-ر) ليتكون مركب فعلي كما هو موضح في القاعدة (٢٠-ز) لتصبح الجملة (S) في النهاية تتكون من مركب اسمي وآخر فعلي.

وبعد اختيار المحلل النحوي لأحد الاتجاهين السابقين، فإن المحلل النحوي عليه أن يحدد ما إذا كان الاتجاه المتبنى للتحليل غير محدد (Nondeterministic Parsing) أو محدد (Deterministic Parsing)؛ حيث تظهر الحاجة إلى التحليل غير المحدد عند ظهور جمل تحتمل أكثر من تحليل، حيث إن القواعد تكون قادرة على أن تسلك أكثر من اتجاه. فهذه الطريقة من التحليل تسمح بإنتاج جميع التراكيب الممكنة في حالة الجمل التي يحدث فيها لبس لغوي، وذلك لتحديد التحليل الأفضل للسياق. ففي هذه الحالة يبدأ المحلل باختيار قاعدة معينة من مجموعة القواعد المحتملة لتركيب الجملة تاركا إشارة تمكنه من معرفة أن في هذا الموضع كان هناك قاعدة أخرى محتملة وبعد ذلك يبدأ مساره في التحليل مع القاعدة المختارة ويستمر في تطبيق القواعد قيد التطبيق التي تمكنه من استكمال مساره في بناء البنية التركيبة للجملة، وعندئذ يتراجع البرنامج (Backtrack) إلى موضع الإشارة التي تركها في بداية مساره الأول لتجربة القاعدة الأخرى البديلة التي تمكنه من استخراج البنية التركيبية الأخرى للجملة؛ فعلى سبيل المثال عند تحليل الجملة في (١٧) نجد أن المحلل أثناء تحليله للبنية التركيبية للجملة ظهر له من خلال القواعد أن الجملة تحتمل أكثر من تحليل؛ حيث إن المركب الاسمي «عقد مؤتمر للسلام في الشرق الأوسط» يمكن تحليل تركيبه وفقا للقاعدة (١٨-أ) فهو بذلك يسلك المسار الذي يحقق التمثيل في الشكل (٤-٢٦) ويمكن أيضا أن يسلك التحليل مسار آخر بادئا بقاعدة أخرى محتملة في (١٨-خ) فهو بذلك يسلك المسار الذي يحقق

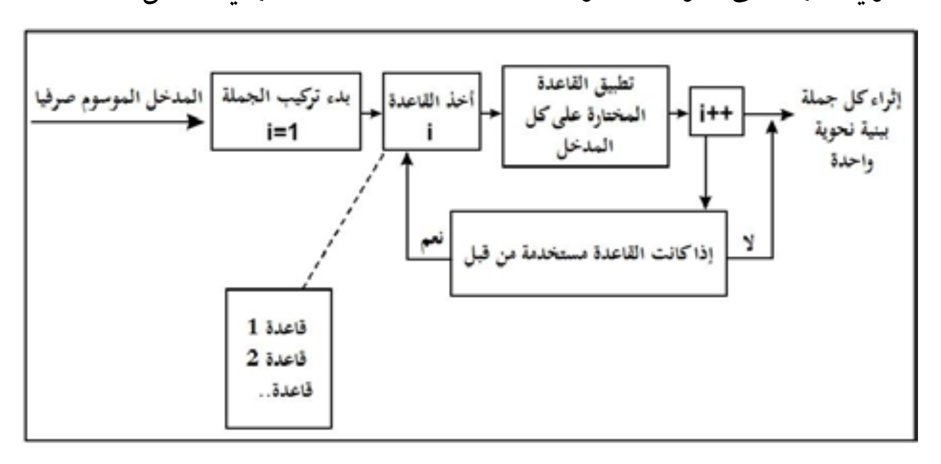
التمثيل في الشكل (٤-٢٧)، وبهذه الطريقة يكون المحلل قادرًا على استخراج أكثر من تركيب نحوي للجملة.

أما إذا كان اتجاه التحليل محدد فإن المحلل ينتقل من تطبيق قاعدة إلى أخرى سالكا مسار واحد فقط لاستخراج بنية تركيبية واحدة فقط للجملة، فلا يضطر البرنامج إلى التراجع إلى نقطة سابقة لاختيار قواعد بديلة أخرى. ولمزيد من التفاصيل عن اتجاهات بناء المحللات النحوية ينصح بالاطلاع على (Grune, & Jacobs, 2007).

٥- اتجاهات بناء المحللات النحوية

٥,١- الاتجاه المبني على القواعد

عند بناء المحلل النحوي بالاعتماد على هذا الاتجاه تتم كتابة معلومات عن البنى النحوية للغة في شكل قواعد نحوية، وهذه هي القواعد التي يتم من خلالها التحليل النحوي للنص المدخل من أجل إنتاج الأشجار النحوية، ويتم عادة تخزين معلومات نحوية عن الكلمات المفردة، وهذه المعلومات تتمثل في نوع الكلمة في المعجم والتي يصل إليها المحلل قبل تطبيق القواعد اللغوية. ويمكن تلخيص خطوات بناء المحلل النحوي اعتمادا على القواعد اللغوية (Kiril Ribarov, 2002) كما في الشكل (٤-٣٠).



الشَّكل ٤-٣٠: خطوات بناء المحلل النحوي المبني على القواعد.

في المخطط الموضح بالشكل (٤-٣٠) يتم وسم المدخل صرفيًا ثم البدء في بناء البنية النحوية للجملة وذلك عن طريق تطبيق عدد من القواعد الموجودة والتي تم الحصول عليها مسبقًا على أساس جمل محللة يدويا. وعلى الرغم من انتشار استخدام المحلل المبني على القواعد في أنظمة المعالجة الآلية للغة الطبيعية، إلا أن هناك العديد من التحديات لبناء هذا المحلل فهو مثلاً يحتاج إلى لقواميس حاسوبية بالإضافة إلى الحاجة إلى لغويين حاسوبيين ذوي مهارة عالية لكتابة ولمراجعة القواعد اللغوية.

٥,٢- الاتجاه المبني على الإحصاء

اعتمد المحللون المعاصرون في السنوات الأخيرة على هذا الاتجاه الإحصائي في بناء المحللات النحوية نظرًا للقدرة المباشرة والسريعة لإخراج النتائج. حيث يتميز ذلك الاتجاه بالقلة الملحوظة للقواعد المطلوبة لبناء المحلل النحوي، وأيضًا القدرة على ضبط المحلل من أجل تحليل نوع معين من النصوص عن طريق استخراج المعلومات الإحصائية. ويعتمد هذا الاتجاه على بناء نموذج إحصائي هدفه تحديد الاحتمالات الإحصائية للتحليلات المختلفة للجمل. وتعتمد هذه الاحتمالات على استخدام عدد ضخم من النصوص الموسومة بمعلومات لغوية كعينات تدريبية. ويعد المحلل النحوي الإحصائي أحد الوسائل الشائعة لحل مشكلة اللبس التركيبي، حيث يقوم النموذج الإحصائي بحساب أكثر التحليلات احتمالا اعتمادًا على العينات التدريبية. ومن أكثر التحديات وضوحًا في هذا الاتجاه اشتراط وجود كميات كبيرة من المدونات الموسومة صرفيًا ونحويًا لقلة وفرتها. ومن الملاحظ أن دقة المحللات الإحصائية الحالية جيدة ولكنها لم ترقى إلى دقة التحليل اليدوي. يمكن الاطلاع على المزيد عن تقنيات التحليل الإحصائي بالرجوع إلى (Bunt, & Nijholt, 2013).

٥,٣- الاتجاه المختلط المبني على القواعد والإحصاء

تواجه المحللات المبنية على القواعد مشكلات وتحديات عديدة مثل اللبس التركيبي، كما أن المحللات الإحصائية تعتمد بشكل أساسي على وجود عينات تدريبية ضخمة من أجل الحصول على نتائج أفضل بشأن احتمالات التحليل التركيبي للجمل. أي أنه يوجد حد أقصى من الكفاءة يمكن تحقيقها عند الاعتماد على اتجاه تحليل دون الآخر (Foth, &

Menzel, 2006). فعلى الرغم من أن المحلل الإحصائي قد يعطي أداء أفضل، إلا أنه يواجه مشكلات عديدة نتيجة عدم دعم قدرته على مراعاة مرونة اللغة والترتيب الحر للكلمات داخل تركيب الجمل (Selvam, & Thangarajan, 2008).

وللتغلب على هذه العيوب في كلا الاتجاهين ظهرت بعض الآراء التي تنادي بتبني دمج القواعد اللغوية مع العمليات الإحصائية. فنجد أن محللات الاتجاه المختلط المبني على القواعد والإحصاء تعتمد على القواعد الخاصة باللغة بالإضافة إلى بعض العمليات الإحصائية التي تعتمد على الاحتمالات التي توضح مدى انتشار القواعد ومدى تطبيقها في الاستخدام اللغوي. فقد وجد أن اللغويون يجيدون كتابة القواعد النحوية الخاصة باللغة ويفضلون الاعتماد على الاتجاه المبني على القواعد ولكن نظرًا لكثرة الغموض التي تسببه القواعد اللغوية كان هناك اتجاه إلى الاعتماد على بعض العمليات الإحصائية التي تساعد في عملية فك الغموض التي تسببه القواعد النحوية. وأصبحت المحللات النحوية المبنية على الاتجاه القواعدي الإحصائي في الآونة الأخيرة أكثر نجاحا إلى حد ما في عمليات التحليل النحوي، ولمزيد من المعلومات عن الاتجاه المختلط المبني على القواعد والإحصاء يمكن الاطلاع على (Foth, & Menzel, 2006).

٦- دقة المحللات النحوية

يمكن النظر إلى المحللات النحوية وتصنيفها من حيث التطبيقات القائمة على هذه المحللات ودرجة تحليل وتفصيل المحلل إلى: محلل نحوي سطحي ومحلل نحوي عميق. فطبيعة بعض التطبيقات الخاصة بالمعالجة الآلية للغات الطبيعية لا تستدعي بناء محلل نحوي عميق وكامل، خاصة تلك التطبيقات التي تعتمد على كمية ضخمة من النصوص المدخلة؛ حيث إن المحلل السطحي يكون كافيًا فقط من أجل الأغراض التي بُنيت من أجلها هذه التطبيقات. فعلى سبيل المثال برامج استرجاع المعلومات تعتمد بشكل أساسي على وجود محلل نحوي سطحي يعمل على استنباط وتحديد المركبات الاسمية والمركبات الفعلية. أما التطبيقات الأخرى مثل الترجمة الآلية القائمة على القواعد فطبيعتها تستوجب وجود محلل نحوي عميق لفك الالتباس اللغوي واستكشاف كل التحليلات الممكنة للجملة. ولسنا هنا بصدد الحديث عن اتجاهين مختلفين للتحليل النحوي يتناقضان في طبيعة المسلك المنهجي لكل منهما، إنما هي خطة عمل يتم توجيهها

حسب الهدف المنشود من بناء المحلل النحوي والتطبيقات الأخرى المستفيدة منه. وكما أن كلا الاتجاهين (العميق والسطحي) يختلفان في منهجية ورؤية التعامل مع النص المدخل، إلا أن الفصل بينهما فصلاً جذريًا يشكل عقبة في طريق التطبيقات المبنية على أساس تحليلي سطحي بحت دون التطرق لآليات التحليل العميق. فعلى سبيل المثال نجد أن أنظمة تحويل النص المكتوب إلى منطوق تقوم في الغالب على أساس محلل نحوي سطحي وظيفته فصل وتقسيم الجمل إلى كتل نحوية ووحدات يمكن الربط بينها داخل الهيكل الداخلي للجملة. ولكن هذه العملية لا تنطبق دائما بسهولة ويسر. ففي بعض الحالات لا يمكن الاكتفاء بآلية التحليل النحوي السطحي بمعزل عن التحليل العميق الذي يقدم معلومات لغوية وحلول أكثر عمقا ودقة. وعليه فإنه لا مانع من الدمج بين النهجين من أجل الوصول إلى رؤية أفضل وتحليل لتركيبات لغوية مختلفة بشكل أكثر فاعلية ومرونة (Balfourier, Blache, & Van Rullen, 2002).

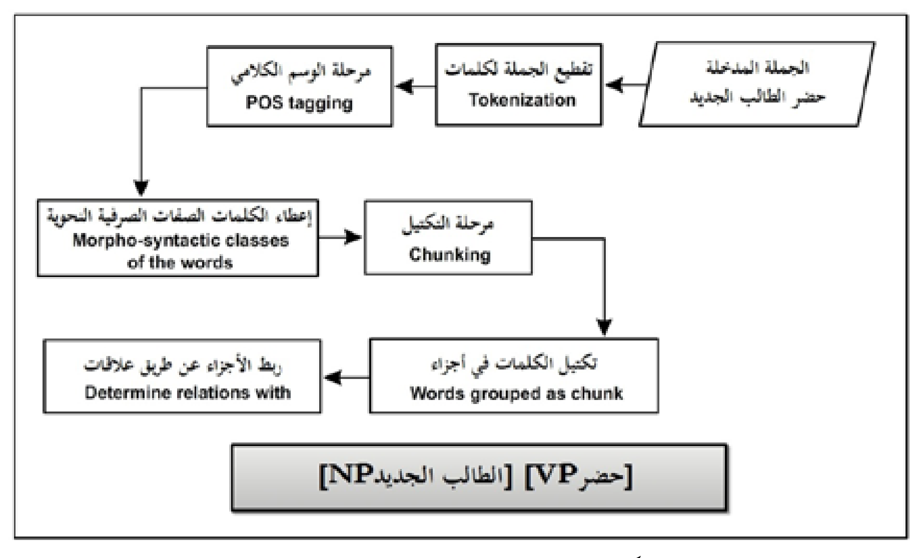
١,٦- التحليل السطحي

ويعرف أيضًا بالتحليل المقداري أو التحليل البسيط أو التحليل الجزئي. ويُعرف التحليل السطحي بأنه تحليل للجمل عن طريق عملية تجزئة النص إلى كتل نحوية؛ أي تقسيم النص إلى كتلٍ، كل كتلة عبارة عن مجموعة كلمات متجاورة تمثل تركيبًا يسهل تمثيله ومعالجته منفصلاً (Balfourier, Blache, & Van Rullen, 2002; Ibrahim, Mahmoud, & El-Reedy, 2016; Mohammed, & Omar, 2011). وهذا النوع من التحليل لا يعتمد على التحليل المفصل لمكونات الجملة، ولا دورها في الجملة الرئيسية. وعليه فإن الهدف الأساسي من المحلل السطحي هو تمثيل الجملة عن طريق عدد محدود من المعلومات اللغوية.

ولقد تعددت اتجاهات التحليل السطحي من حيث اعتمادها على القواعد اللغوية داخل بنية الجملة ومن حيث اعتمادها على الاحتمالات والتكرارات الإحصائية داخل النص. فنجد أن الاتجاه المبني على القواعد يعتمد على استخدام عدد من المحولات محدودة الحالات (Finite State Transducers) لتقوم بمهمة تحديد الكتل الاسمية والكتل الفعلية إلخ بناء على القواعد المخزنة. أما الاتجاه المبني على الإحصاء يتم فيه تحديد الكتل النحوية عن طريق التعرض لكمية مناسبة من البيانات التدريبية تم وسمها

من قبل وتحديد الكتل اللغوية فيها. وتتميز التطبيقات المبنية على آلية التحليل السطحي بأنها تتعامل مع كمية ضخمة من النصوص المدخلة؛ فعلى سبيل المثال برامج استرجاع المعلومات وبرامج استخراج المعلومات وبرامج التلخيص الآلي تعتمد بشكل أساسي على وجود محلل نحوي سطحي يعمل على استنباط وتحديد المركبات الاسمية والمركبات الفعلية من الجمل المدخلة. وبرامج الإجابة عن الأسئلة أيضًا تقوم على تحديد العلاقات النحو- دلالية داخل بنية السؤال لاستنباط تراكيب بعينها كتحديد الفاعل والمفعول به والمكان والزمان إلخ.

يمكن وصف التحليل السطحي بشكل مبسط كما هو في الشكل (٤-٣١) حيث يمر المحلل بعدد من المراحل. أولا مرحلة الإعداد وتشخيص النص إلى كلمات. وهذه مرحلة أولية تهدف إلى إعداد النص وذلك عن طريق تقسيم وفصل النص إلى كلمات. ثم تأتي مرحلة تصنيف أقسام الكلام حيث يتم في هذه المرحلة تصنيف كل كلمة من كلمات الجملة حسب نوع كل كلمة وذلك لا يتم بمعزل عن المعجم. وبعد ذلك تأتي مرحلة التكتيل أي تقسيم النص إلى كتل كل كتلة عبارة عن مجموعة كلمات متجاورة تمثل تركيب، كل تركيب يسهل تمثيله ومعالجته. وأخيرًا ننتهي بالعلاقات النحوية وهي مرحلة ربط الكتل النحوية بعلاقات تربطها داخل هيكل الجملة.



الشَّكل ٤-٣١: مراحل التحليل السطحي.

فعلى سبيل المثال جملة (ذهب الطبيب الجديد إلى المستشفى) يتم تحليلها كما في الشكل (٤-٣٢):

[ذهب VP] [الطبيب الجديد NP] [إلى المستشفى PP]

الشَّكل ٤-٣٢: التحليل السطحي لجملة «ذهب الطبيب الجديد إلى المستشفى».

ومن أمثلة المحللات السطحية: محلل كولينز (Collins Parser)؛ هو محلل سطحي إحصائي تم تطويره بدايةً كأطروحة للدكتوراه في جامعة بنسلفانيا. وهذا المحلل لا يستخدم فقط من أجل تحليل الجمل نحويا، ولكن أيضًا يمكن استخدامه كأحد المكونات لتطبيقات أخرى مثل برامج التلخيص الآلي، وبرامج استخراج المعلومات وبرامج الترجمة الآلية وبرامج الإجابة على الأسئلة (M. Kaplan, R., et al., 2004).

ومن الخصائص التي تميز هذا المحلل أن الفئات اللغوية في الجمل المحللة يمكنها التمييز فيها بين رأس الجملة، والمكملات، والملحقات. وهذه الفروق يمكن أن توجه في بعض التطبيقات لاستخلاص علاقات دلالية هامة.

١,٦- التحليل العميق

ويقصد بالتحليل النحوي العميق أنه على المحلل النحوي الآلي توفير جميع المعطيات اللازمة للتحليل النحوي الأعمق. والحديث عن العمق هنا ليس حديث عن النهج المبني على القواعد في مقابل النهج المبني على الإحصاء. فالاتجاه نحو التحليل العميق يستفيد من كلا المنهجين من أجل الوصول إلى أقصى معلومات لغوية نحوية يمكن استخلاصها من الجمل. وأحد أهداف المحلل العميق القدرة على اكتشاف وتحليل كل التراكيب النحوية التي ينتجها الإنسان، والقابلية للتعامل مع كل الظواهر اللغوية؛ القياسية منها وغير القياسية. والجدير بالذكر أن المحللات العميقة لا تهدف إلى تقديم تحليل نحوي كامل للجمل، كما قد يعتقد القارئ من المعنى الحرفي للعمق. وإنما الهدف الأساسي من التحليل العميق هو السعي إلى حل التراكيب النحوية التي تنطوي على درجات متفاوتة من العمق والتعقيد، وهذا يتوقف على مدى التعقيد ووجود المعلومات اللغوية الأساسية اللازمة لحل مثل هذه التراكيب (Chanod, 2001).

وتواجه التطبيقات اللغوية القائمة على آلية التحليل العميق بعض العقبات التي تعوق عملها، فعلى الرغم من العمق والثراء اللغوي الذي قد يقدمه المحلل النحوي العميق، إلا أن تطبيقه يشكل عبئًا مقارنة بالمحلل السطحي. فنجد أن المحلل السطحي قد يكون اتجاهًا أفضل من الناحية التطبيقية لعدة عوامل. منها عامل السرعة والوقت المستهلك في التحليل، فالمحلل السطحي يتعامل مع الجمل من حيث تقسيمها إلى كتل أصغر دون الخوض في تفاصيل ومشاكل أخرى كفك الالتباس والتعامل مع المحذوفات. وهذا بالطبع يأتي على حساب العامل الوقتي للمحلل النحوي. وكذلك فإن طبيعة المحلل السطحي وكونه يتعامل مع كتل لغوية قصيرة تم فصلها من الجملة؛ فإن هذا يتطلب مساحة أقل من ذاكرة التخزين ويؤثر على سرعة وعملية التحليل (Collins,1996). كما أنه في حالة إذا أخفق المحلل العميق في تحليل جملة ما، يقوم بإسقاط الجملة بكاملها نتيجة لنظرته الكلية للمدخلات. إما في حالة المحلل السطحي فإذا أخفق في تحليل جملة فإنه يقوم بإسقاط الكلمات التي يصعب التعامل معها وتكتيلها في كتل محددة دون اللجوء إلى إسقاط الجملة كاملةً. ويمكننا تقديم مثال للتحليل العميق؛ الجملة (الولد ذهب بسرعة إلى البيت) يتم تحليلها كما في الشكل (٤-٣٣):

S[VP[PP[NP[N[بيت] DET[ال]P[إلى] ADV[بسرعة]V[ذهب] NP[N[ولد]DET[ال]

الشَّكل ٤-٣٣: التحليل العميق لجملة «الولد ذهب بسرعة إلى البيت».

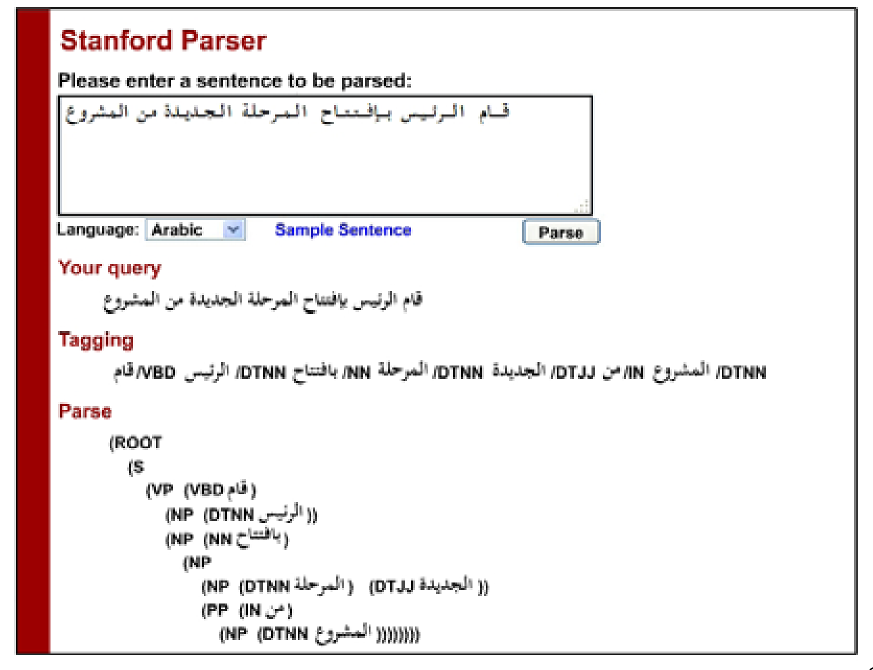
٧- أمثلة للمحللات النحوية

الآن وبعد عرض بعض الاتجاهات المتبعة عند بناء المحلل النحوي الآلي، سنتعرف على بعض أمثلة المحللات الآلية الحالية للعربية. ووجب التنبيه على أن العربية لازالت تعاني من نقص في المحللات الآلية والأدوات اللازمة لبنائها. ومن المحللات الحالية محلل ستانفورد الإحصائي والمحلل التفاعلي (IAN) وهو ما سنقوم بشرح آليات تحليله للجملة العربية بالتفصيل.

١,٧- محلل ستانفورد الإحصائي

وهو محلل إحصائي يستخدم المعرفة المكتسبة من جمل محللة يدويا في محاولة لإنتاج التحليل الأرجح لجمل جديدة. وهذا المحلل يعمل على اللغة الانجليزية ولغات أخرى منها الإيطالية والبرتغالية والبلغارية والعربية.

ومحلل ستانفورد الإحصائي يستخدم تقديرات احتمال الأرجحيَّة القُصوى (Maximum Likelihood Estimates) لحساب الاحتمالات من القواعد النحوية. وهناك مصدران أساسيان للخطأ في التقدير ((O'Reilly, 2010، أولا التحيز الذي يكون احتمال الأرجحيَّة القُصوى له صفر. وثانيا الخطأ في اختيار العينات، والذي يرجع لمشكلة ضخامة البيانات المدخلة بسبب العدد الكبير من القواعد النحوية. ولذلك فإنه من المقبول عمومًا أن يقدم المحلل شكل متجانس يمكن أن يساعد في تحسين عمومية تقديرات احتمال الأرجحيَّة القُصوى في هذه الحالة وتحسين الأداء. وفي الشكل (٤-٣٤) مثال لتحليل جملة «قام الرئيس بافتتاح المرحلة الجديدة من المشروع» داخل محلل ستانفورد:



الشَّكل ٤-٣٤: مثال لتحليل جملة «قام الرئيس بافتتاح المرحلة الجديدة من المشروع» داخل محلل ستانفورد.

٢,٧- المحلل التفاعلي (Interactive analyzer- IAN)

قامت منظمة لغة التواصل الرقمية العالمية (UNDL) بوضع أدوات لمساعدة اللغويين في إنتاج الموارد اللغوية المختلفة. ومن هذه الأدوات التي وضعتها المنظمة أداة المحلل التفاعلي (IAN) لتحليل اللغات الطبيعية. وهو محلل مبني على نظرية X–bar وهي نظرية لسانية توليدية من النظريات اللغوية،كما تم شرحها من قبل في جزء المركبات الوسيطة (Alansary, Nagi, & Adly, 2010).

والمحلل النحوي التفاعلي يتبنى اتجاه التحليل النحوي من أسفل إلى أعلى. ولتوضيح كيفية حدوث التحليل الآلي النحوي في المحلل التفاعلي سيتم شرح المراحل التي يجب استخدامها لنقل جملة عربية إلى شجرتها النحوية عن طريق الجملة في (٢١):

أكل الولد تفاحة لذيذة (٢١)

أولاً تمر الجملة على القاموس الذي يحتوي على مفردات اللغة وبعض الصفات التي توضح المعلومات اللغوية الخاصة بتلك المفردات كالمعلومات المورفولوجية التي تدل على الوسم الخاص بها، ولمزيد من التفاصيل يمكنك الاطلاع على الرابط:

http: //www.unlweb.net/wiki/index.php/Part_of_speech.

ففي حالة «أكل الولد تفاحة لذيذة» فإن المحلل النحوي سوف يقوم باستخراج كل المفردات التي تطابق كلمات الجملة من القاموس كما في الشكل (٤-٣٥).

[أكل] (V)
[أكل] (N)
[ال] (DET)
[ولد] (N)
[تفاحة] (N)
[لذيذة] (J)

الشّكل ٤-٣٥: نموذج من القاموس يمثل مفردات الجملة في (٢١).

ففي «أكل» يوجد هناك غموض حيث يوجد «أكل» بوسم اسم و «أكل» بوسم فعل. ولكن في هذه الجملة يجب اختيار «أكل» بوسم فعل وليس اسم. وبالتالي فإذا

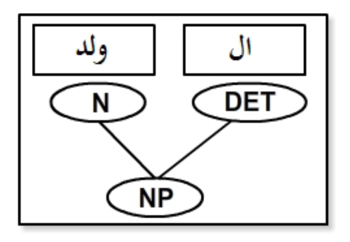
كانت القواعد قادرة على اختيار الفعل بدلا من الاسم فهذا سيسهل عملية التحليل. ولكن في حالة اختيار «أكل» بوسم اسم فيأتي دور صفة التفاعلية في المحلل. لأن المحلل يتيح للمستخدم فرصة اختيار المفردات المناسبة يدويًا قبل مرحلة التحليل النحوي.

وبعد الانتهاء من اختيار المدخلات المعجمية المناسبة لمفردات الجملة العربية فيبدأ المحلل بحذف المسافات بين مفردات الجملة المدخلة لتصبح شكل الجملة المدخلة كما في الشكل (٤-٣٦):

["أكل"]["ال"]["ولد"]["تفاحة"]["لذيذة"]

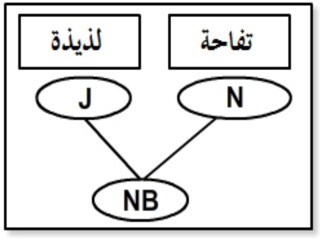
الشَّكل ٤-٣٦: تحديد قائمة مفردات الجملة.

ثم يبدأ المحلل النحوي بناء المركبات النحوية المختلفة؛ فعند تجاور أداة التعريف «ال» - التي تُعتبر المخصص النحوي الاسمي - والاسم «ولد» يُبنى المركب الاسمي النهائي كما في الشكل (٤-٣٧):



الشَّكل ٤-٣٧: بناء المركب الاسمي (NP).

وعند تجاور الاسم «تفاحة» والصفة «لذيذة» يتم بناء المركب الاسمي الوسيط كما في الشكل (٤-٣٨):



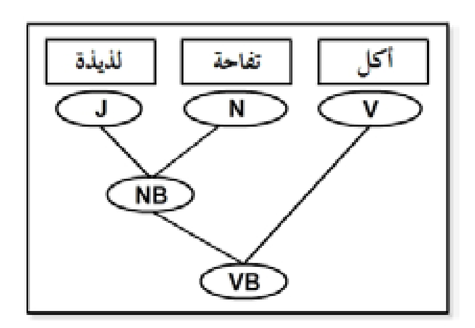
الشَّكل ٤-٣٨: بناء المركب الاسمي الوسيط (NB).

ثم تأتي القاعدة النحوية التي تنص على أنه عند تجاور فعل ومركب اسمي ومركب اسمي آخر فسوف يتم تبديل مكان المركب الاسمي الأول «الولد» ليسبق الفعل لكي يكون هذا المركب موجود في موقع المخصص النحوي للفعل في الشجرة النحوية وهكذا يكون المركب الاسمي الثاني «تفاحة لذيذة» في موقع المتمم النحوي للفعل، كما في الشكل (٤-٣٩):

NP [["ال"] ["ولد"]] ["أكل"] NB[["تفاحة"] ["لذيذة"]]

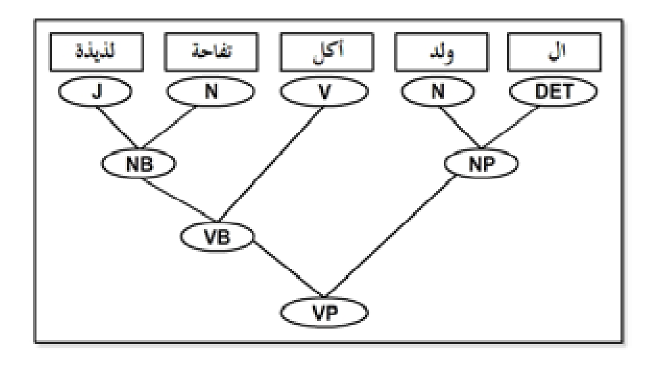
الشَّكل ٤-٣٩: تجاور المركب الاسمي الوسيط « تفاحة لذيذة» بجانب الفعل «أكل».

وهذا المركب الاسمي الوسيط الذي تم بناؤه في الشكل (٤-٣٨) يمثل المفعول به والذي يتم ربطه بالفعل «أكل» وبالتالي نصل إلى بناء المركب الفعلي الوسيط كما في الشكل (٤-٤٠):



الشَّكل ٤-٤٠: تكوين المركب الفعلي الوسيط «أكل تفاحة لذيذة».

ولاستكمال بناء الشجرة النحوية يأتي دور القاعدة النحوية التي تربط المركب الاسمي النهائي «الولد» والتي تمثل الفاعل مع المركب الفعلي الوسيط لتكوين المركب الفعلي النهائي كما في الشكل (٤-٤١).



الشَّكل ٤-٤١: تمثيل التركيب النحوي للجملة في (٢١).

وهكذا يكون قد تم بناء الشجرة النحوية التي تمثل التركيب النحوي للجملة في (٢١) ويمكننا أيضا كتابة التمثيل النحوي للجملة في (٢١) كما في الشكل (٤-٤٢) خلال المحلل التفاعلي.

```
[S:1]
 {org}
أكل الولد تفاحة لذيذة
 {/org}
 {unl}
VP(:03,      :01)
VB:03(أكل   ,     :02)
NB:02(لذيذة  ,        تفاحة)
NP:01(ال  ,          ولد)
{/unl}
 [/S]
```

الشَّكل ٤-٤٢: تمثيل التركيب النحوي للجملة في (٢١) باستخدام المحلل التفاعلي.

ونجد أن المحلل يقوم بوضع أرقام أمام الفروع المكونة للشجرة لربطها داخل هيكل الشجرة النحوية؛ فالمركب الاسمي الوسيط «تفاحة لذيذة» الذي يمثل المكمل النحوي للفعل يسمى في الشجرة التي تنتج من المحلل ب «٠٢» ويُربط هذا المركب مع

الفعل «أكل» ليكون المركب الفعلي الوسيط الذي يسمى ب «٠٣». ثم المركب الاسمي النهائي «الولد» الذي يمثل المخصص النحوي للفعل ويسمى ب «٠١» الذي يُربط بالمركب الفعلي الوسيط الذي يسمى «٠٣» ليكون المركب الفعلي النهائي «أكل الولد تفاحة لذيذة». وهكذا نجد أن التمثيل النحوي في الشكل (٤-٤٢) يطابق التمثيل الشجري في الشكل (٤-٤١).

٨- بعض النقاط البحثية الهامة من أجل دعم خطة طريق لمعالجة التركيب في اللغة العربية

أ. دراسة القواعد المنظمة لتتابع الأسماء داخل المركبات الاسمية.

ب. دراسة إمكانية تقسيم الكلمات العربية بشكل يدعم معالجة التركيب النحوي.

ج. دراسة السمات اللغوية المختلفة للكلمات للمساعدة في ضبط وانتظام تطبيق القواعد.

د. دراسة الإطار النحوي لمفردات اللغة العربية المختلفة وذلك لإثراء المعجم بمزيد من المعلومات النحوية والتي لها فوائد جمة تدعم التحليل النحوي للجمل العربية.

هـ. دراسة قيود الانتقاء (Selection Constraints) لمفردات العربية دعما للمعجم ولدوره الجليل أثناء التحليل النحوي.

و. دراسة إمكانية الوصول لقواعد تحديد حدود المركبات المختلقة داخل الجملة.

ز. التحليل المنهجي لعلاقة منظومة النحو العربي بمنظومتي الصرف والدلالة.

ح. إعادة النظر في جهود النحاة العرب القدامى من أجل تنظير أفضل لربط النظريات الحديثة بالجهود القديمة.

ط. تفعيل استخدام النظريات المختلفة للغات الأخرى، مع ضرورة الأخذ في الاعتبار طبيعة اللغة العربية.

ي. تطبيق نظرية X-bar على التراكيب المختلفة في الجمل العربية.

ببليوجرافيا مرجعيَّة

١. جون سيرل، تشومسكي والثورة اللغوية، مجلة الفكر العربي، عدد ٨-٩، ص١٢٦، ١٩٧٩.

٢. ميشال زكريا، الألسنية والتوليدية والتحويلية وقواعد اللغة العربية، المؤسسة الجامعية للدراسات والنشر والتوزيع، ط٢، ١٩٨٦.

٣. نبيل علي، اللغة العربية والحاسوب، دار تعريب، ١٩٨٨.

4. Al-Ansary, S., & El-Kareh, S. (2004, April). Arabic-English Machine Translation Systems: Discrepancies and Implications. In JEP/TALN International Conference, Special session on Arabic text and speech language processing.

5. Alansary, S., Nagi, M., & Adly, N. (2010, December). UNL+ 3: The gateway to a fully operational UNL system. In 10th International Conference on Language Engineering, Ain Shams University, Cairo, Egypt.

6. Balfourier, J. M., Blache, P., & Van Rullen, T. (2002, August). From shallow to deep parsing using constraint satisfaction. In Proceedings of the 19th international conference on Computational linguistics-Volume 1 (pp. 1-7). Association for Computational Linguistics.

7. Bunt, H., & Nijholt, A. (Eds.). (2013). Advances in probabilistic and other parsing technologies (Vol. 16). Springer Science & Business Media.

8. Chanod, J. P. (2001). Robust parsing and beyond. In Robustness in Language and Speech Technology (pp. 187-204). Springer, Dordrecht.

9. Chomsky, N. (1956). Three models for the description of language. IRE Transactions on information theory, 2(3), 113-124.

10. Collins, M. J. (1996, June). A new statistical parser based on bigram lexical dependencies. In Proceedings of the 34th annual meeting on Association for Computational Linguistics (pp. 184-191). Association for Computational Linguistics.

11. Foth, K. A., & Menzel, W. (2006, July). Hybrid parsing: Using probabilistic models as predictors for a symbolic parser. In Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics (pp. 321-328). Association for Computational Linguistics.

12. Grune, D., & Jacobs, C. J. (2007). Parsing Techniques. Monographs in Computer Science. Springer,, 13.

13. Ibrahim, M. N., Mahmoud, M. N., & El-Reedy, D. A. (2016). Bel-Arabi: advanced Arabic grammar analyzer. International Journal of Social Science and Humanity, 6(5), 341.

14. Kaplan, R., Riezler, S., King, T. H., Maxwell III, J. T., Vasserman, A., & Crouch, R. (2004). Speed and accuracy in shallow and deep stochastic parsing. In Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics: HLT-NAACL 2004. Katamba, F. & Stonham, J. (1993). Morphology, Palgrave Modern Linguistics.

15. Katamba, F. (1993). Morphology. New York: St.

16. Marimon, M., & Bel, N. (2004). Lexical Entry Templates for Robust Deep Parsing. In LREC.

17. Miyao, Y. (2006). From linguistic theory to syntactic analysis: Corpus-oriented grammar development and feature forest model. PhD thesis, University of Tokyo.

18. Mohammed, M. A., & Omar, N. (2011). Rule based shallow parser for Arabic language. Journal of Computer Science, 7(10), 1505-1514.

19. O'Reilly, C. (2010). These go to eleven'Investigations in tuning the Stanford Statistical Parser.

20. Dobrovolsky, M., Katamba, F., & O'Grady, W. D. (Eds.). (1997). Contemporary linguistics: an introduction. St. Martin's Press.

21. Radford, A. (1988). Transformational grammar: A first course (Vol. 1). Cambridge University Press.

22. Selvam, M., Natarajan, A. M., & Thangarajan, R. (2008). Structural parsing of natural language text in tamil using phrase structure hybrid language model. International Journal of Computer, Information and Systems Science, and Engineering, 2008, 2-4.

الفصل الخامس

التَّحليل الدِّلاليّ

د. مُحمَّد عَطيَّة

١- الأنطولوجيا ودِلالة اللغة

٢- التحليل الدِّلالي المُعْجَمي.

٣- المُعالَجَة الدِّلالية المُعْجَمية في اللغة العربية.

٤- شبكات الدِّلالات المُعْجَمية كإطارٍ أُنْطُولوجيٍّ جزئي.

٥- الالتباس الدِّلالي والعمل على إزالته.

٦- تطبيقات التحليل الدِّلالي المُعْجَمي.

٧- العَنْونة الدِّلالية المُعْجَمية للمدوَّنات النصية العربية.

٨- التحليل الدِّلالي ما بعد المستوى المُعْجَمي.

١ - الأنطولوجيا ودِلالة اللغة

يعود الانشغال بالدراسات الدِّلالية إلى أقدم عصور الحضارات الإنسانية في التاريخ البشري المدوَّن، حيث انشغل العلماء والمفكرون والفلاسفة ورجال الدين بتأمل العالم وكائناته وظواهره واجتهد كلٌّ منهم في محاولة تفسير معانيها ونَظْمِها سويًّا في سياق عام وفْقَ أنساق منتظمة.

وربما كان أحد أنضج المفاهيم النظامية التي تبلورت عبر الزمن لتعبِّر عن هذه التساؤلات الوُجودية هو ما يُعَبَّر عنه بمصطلح «الأُنْطُولُوجْيَا» (Ontology) - وهو اسمٌ مشتَقٌّ من لفظة «أُنْطو» اليونانية القديمة التي تعني حَرْفيًّا «الوجود أو الكينونة» - ويتبوأ المصطلح مكانته في الفلسفة كعنوان على أحد مباحثها الرئيسية ويُعْنَى بدراسة طبيعة الوجود، وتصنيف الموجودات/ الكائنات والعلاقات بينها، والتمييز بين العام والخاص، وبين الكليات والجزئيات، وبين ما هو جوهرٌ وما هو كائنٌ، وبين الخصائص الذاتية والصفات الخارجية [٢٠].

ومن الوجهات الشائعة للبحث الأنطولوجيّ طموحًا إلى بناء نموذج مَعْرِفيّ للعالَم هو العمل على الوصول إلى تقسيم طبيعي لكل الجُزْئيَّات المكوِّنة له عند مَفاصِلِه الأساسية بغرض اكتشاف الفئات التي تندرج تحتها موجوداته. وفي حين أسهب الفلاسفة في جدل لا يكاد ينتهي فيما بينهم منذ منتصف القرن العشرين الميلادي حول التعريفات والمقارَبات والطرق المثلى لبناء أنطولوجيات مكتملة ومتماسكة، فإنهم لم يقوموا في حقيقة الأمر بإنجاز أية أنطولوجيا مكتملة ومفصَّلة يُعْتَدُّ بها!

أما ما يهمنا من ناحية عملية في هذا الفصل من الكتاب فهو ما قام به الباحثون في مجال علوم الحاسبات عمومًا وفي حقل الذكاء الاصطناعي على وجه الخصوص، حيث التقطوا منذ منتصف سبعينات القرن العشرين الميلادي أهمية مفهوم الأنطولوجيا واستلهموه إطارًا لجمع المعرفة وتمثيلها بشكل منهجي منظم يمكن الاستفادة منه حاسوبيًّا. [٢١] وقد تواكب هذا الاستلهام مع انتهاء باحثي الذكاء الاصطناعي إلى أن «الخُوارِزْمات الحاسوبية»[١] (Algorithms) الحاذقة (ومن أمثلة هذه الخوارزمات

١ - نُذكِّر هنا بأن الخُوارِزْم الحاسوبي هو عبارة عن الوصف الرياضي المفصل للخطوات الحاسوبية التي يؤدي تنفيذها لحل إحدى المسائل الرياضية العامة.

«أساليب البحـــث الشَّجريَّة Tree Search Methods») لا تكفي وَحْدَها لحل المسائل الواقعية الجادة، بل لا بد كذلك من مزاوجتها بقاعدة معرفية حسنة التمثيل للمسألة المطلوب التعامل معها، وبذلك فإن خوارزمات المعالجة الحاسوبية الحاذقة بالتزاوج مع القواعد المعرفية المحوسَبة (الأنطولوجيات) تشكلان القدمين اللتين يمشي عليهما الذكاء الحاسوبي القادر على التعامل مع مسائل صعبة كانت فيما مضى لا تستغني في مُقارَبتها عن الذكاء البشري.

ومع وجود بعض الاختلافات الفرعية بين الباحثين والمشتغلين بعلوم الحاسب فيما يتعلق بالبِنَى الأنطولوجية، فإن هناك اتفاقًا عامًّا على مُركِّباتٍ أساسية لبِنْيَة الأنطولوجيا؛ ومن أهمها:

أ. الموجودات/ الكائنات وأفراد هذه الموجودات[1].

ب. الفئات/ الأنواع، وتصنيف الموجودات تحت هذه الفئات.

ج. العلاقات بين الموجودات (وكذلك العلاقات بين أفرادها).

د. خصائصُ الفئات (والموجوداتِ تحتَها).

هـ. الأحداث؛ حيث يشير الحدث إلى تغيرٍ في خصائص الأفراد و/ أو الموجودات و/ أو الفئات، و/ أو العلاقات بين أيٍّ منها.

و. الوظائف؛ وهي التي تسمح بتكوين بِنًى مركَّبة من الأفراد يمكن التعامل معها لاحقًا كوحدات أنطولوجية.

ز. القيود؛ وتوضع على ما يمكن قبوله كمدخَلات من صيغ تقريرية لحالة الأنطولوجيا أو جزء منها.

ح. القواعد (الشرطية غالبًا – في شكل مقدمات ونتائج) التي تصف الاستدلال المنطقي الذي يمكن استشفافه من مجموعة الصيغ التقريرية المقبولة.

ط. المسلَّمات.

---

١ – «الإنسان» كائن؛ أما الشخص المسمَّى «فلان بن فلان بن فلان» على سبيل المثال فهو فرد من أفراد هذا الكائن، و«الشارع» كائن؛ أما شارع «الشانْزِلزِيهْ» في باريس عاصمة فرنسا على سبيل المثال فهو فرد من أفراد هذا الكائن، ... إلخ.

وفي حين أن المركِّبات المذكورة أعلاه من «أ» إلى «د» هي مركِّبات ثابتة (اسْتاتِيكيَّة) في زمن التشغيل، فإن المركِّبات «هـ» مع «و» هي مركِّبات متغيِّرة (دَيْناميكِيَّة) في زمن التشغيل، كما أن المركِّبات من «ز» إلى «ط» هي مركِّبات استاتيكية تمهِّد الأرضية لديناميكية الأنطولوجيا.

وبينما يشيع اليومَ بناء وتوظيف أنطولوجيات على درجة عالية من التفصيل في عدد من المجالات الحاسوبية المتقدمة – ومن أمثلة ذلك؛ المحاكاة والألعاب – إلى الحد الذي يعتبر معه البعض «هندسة الأنطولوجيات» فرعًا قائمًا بذاته من «هندسة المعرفة» ، فإن حقل حوسبة اللغات الحية يولي اهتمامًا خاصًّا بالأنطولوجيا.

وقد نشأ الاهتمام المكثَّف من قبل الباحثين في مجال معالجة اللغات الحية بالأنطولوجيا على أساس أنها تستطيع تشكيل الإطار الرابط بين الطبقات الأولية لمعالجة اللغة الحية؛ كالتحليل الصوتيّ (الفُونُولُوجيّ) والصرفيّ (المورْفُولُوجيّ) على سبيل المثال، وبين الطبقات العليا لمعالجة اللغة؛ مثل «الترابط الخطابي» (Discourse Integration) ومعرفة «المغزى السياقي للكلام» (Pragmatics) وهي التي تستلزم معرفةً بواقع العالم وتفاعلاته. ففي غياب إطار معرفي يربط مفردات اللغة وقواعدها الشكلية مع حقائق ووقائع العالم الذي تعبر عنه بما فيه من كائنات وعلاقات بين هذه الكائنات ... إلخ، فإن معالجات اللغة الحية تبقى منحصرة في الإطار الشكلي الذي يقتصر على أنواع من المعالجة الرمزية لا ترتبط بما تدل عليه هذه الرموز في واقع استخدامها [١٩] (سواء كان واقع العالم المشترك الجامع لعموم البشر، أم كان واقعَ عالَمٍ جزئي متخصص منه، أم كان حتى واقعَ عالَمٍ متخيَّل)، وعندئذٍ يحد هذا الغياب كثيرًا من دقة وامتداد ومردود معالجات أية لغة حية.

وعلى ذلك فقد اتجهت حوسبة اللغات الحية منذ ثلاثة عقود أو يزيد إلى اعتماد الأنطولوجيا إطارًا نموذجيًّا لجمع وتمثيل المعرفة اللغوية الدلالية بحيث تَجْسِر بين ما دونها وما فوقها من طبقاتٍ لمعالجة اللغة الحية. ونظرًا لأن بناء أنطولوجيا كاملة بكل مركِّباتها – المشار إليها عاليَهُ – وذات محتوى يشمل معرفتنا عن العالم كان وما يزال طموحًا شاهقًا يمثل إنجازه مَعْلَمًا أساسيًّا على الطريق الطويلة لحوسبة اللغات الحية وتطبيقاتها، فإنه لم يتم حتى الآن بناءُ مِثْلِ هذه الأنطولوجيا الكاملة الشاملة لأية لغة حية وهناك جدل بين العلماء عما إذا كان في المستطاع إنجاز ذلك في

المستقبل المنظور[1]. وقد أدى ذلك بالعلماء والباحثين العاملين على حوسبة اللغة إلى سلوك مَنْحًى عَمَلِيٍّ عبر الاستعاضة عن بناء أنطولوجيا شاملة لِلُّغة ببناء أطر معرفية دِلالية لِلُّغة بعدة أشكال مختصرة تستلهم روح الأنطولوجيا وتحافظ على بعض أهم مركباتها للتمثيل الدلالي - وخصوصًا تلك المرقَّمة عالِيَهُ من أ إلى د - ويمكننا النظر إلى مثل هذه الأطر كأنطولوجيات جزئية سواء من حيث مركِّباتها أو محتوياتها.

٢- التحليل الدِّلالي المعْجَمي

وربما تُعَدُّ أبرز الأطر الأنطولوجية الجزئية التي استقر بناؤها بنجاح لحوسبة المعرفة اللغوية الدِّلالية هي تلك التي تحتوي «الدِّلالات المعجمية» (Lexical Semantics). وكما يُوحِي اسْمُها فإن الدِّلالات المعجمية لا تُغَطِّي كامِلَ طبقة المعالجة الدلالية لِلُّغة الحية ولكنها فقط تشمل المعالجة الدِّلالية لمفرداتها وهي مجرد الشريحة الدنيا من هذه الطبقة، أما المعالجة الدلالية لما فوق ذلك من تراكيب لغوية (عبارات، جُمَل، فقرات ...) فإن محاولات مُقارَبتها حاسوبيًّا عبر أنطولوجيات أكثر توسُّعًا وتعمُّقًا ما زالت في مراحل أبكر بكثير، ويعرض الفصلُ التالي من الكتاب «معالجة الجملة العربية دِلاليا بين التحليل والتوليد» لمحاولاتٍ رائدةٍ لمعالجةِ اللغة العربية دِلاليًّا فيما يتجاوز المستوى المعجمي.

وفيما نُفْرِد المجال في القسم الرابع من هذا الفصل لتأصيل الكِيانات الدِّلالية المعْجَمية كأنطولوجيات جزئية وفق العرض الذي قدمناه في القسم السابق (الأول)، فإن ما تبقى من هذا القسم (الثاني) يَعْرِضُ لِبنْيَةِ هذه الكِيانات والاعتبارات العملية لإنشائها، كما يتعرض القسم التالي (الثالث) لأمثلةٍ من الكِيانات الدِّلالية المعْجَمية العربية القائمة بالفعل وبعض ما يترتب عليها من معالجات دِلالية مُعْجَمية.

ونستطيع أن نرسم صورة دقيقة للدِّلالات المعْجَمية في لُغَةٍ ما كشبكةٍ هائلة تَرْبِط كل مفردات اللغة ببعضها البعض عبر علاقات دِلالية، كما يمكن أن نُشَبِّه هذه الشبكة بنسيجٍ يُمْكِن صياغة خَيْطِهِ الأساسي على النحو التالي:

١- يحتاج إنجاز مثل هذه الأنطولوجيا الكاملة الشاملة مواردَ بشريةً وإداريةً وحاسوبيةً وماليةً هائلة، ولا يُعتقَد أنه من الممكن إنجاز مشروع كهذا في زمن مقبول دون تعاون على مستوى عالمي - مثل ما اتُّبِعَ في مشروع «الجينوم Genome» - ونرجو أن تبادر إحدى المنظمات الإقليمية أو الدولية في أقرب وقت إلى القيام على التوعية بأهميته ومن ثَمَّ الترويج لتمويله وتنفيذه.

ل	*R*
ف$_{١}$ ⟵ <u>ف</u>$_{٢}$	$w_1 \longrightarrow \underline{w}_2$

أي أن المفرَدة ف$_{١}$ ترتبط بالعلاقة ل مع المفرَدة (أو المفرَدات) <u>ف</u>$_{٢}$

وبمعرفة ذلك يَسْهُل تلخيصُ وصياغةُ الوظيفة الأساسية للمعالجة الدِّلالية المعْجَمية في إيجاد الإجابات الصحيحة على السؤالين التاليين:

أ. بمعرفة مُفْرَدتين؛ ما هي العلاقة الدِّلالية التي تربطهما؟ أي أنَّ ف١ مع ف٢ معلومتان والمطلوب هو الحصول على ل.

ب. بمعرفة مُفْرَدة وعلاقة دِلالية؛ ما هي المفردات التي ترتبط بهذه المفردة الأولى عَبْرَ هذه العلاقة الدِّلالية؟ أي أن ف$_{١}$ مع ل معلومتان والمطلوب الحصول على <u>ف</u>$_{٢}$.

تشكِّل إذن كلٌّ من «مُفرَدات اللُّغة» و «العلاقات الدِّلالية» المكوِّنين الرئيسيين اللَّذَيْن تدور حولهما المعالجة الدِّلالية المعجمية. وفي حين أن المفرَدات العربية قد سبق تناولُها بإسهاب في باب «التحليل الصرفي الآلي لمفردات اللغة العربية» من هذا الكتاب، فإن ما نحتاج لإضافته بشأنها في هذا المقام هو افتراض أنها فئة مغلقة ذات حجم = م مفردةً. أما العلاقات الدِّلالية فهي كذلك فئة مغلقة معرَّفةٌ سَلَفًا ذات حجم = ع نوعًا من الروابط التي يصل كلٌّ منها بين بعض الأزواج المرتَّبة من مفردات اللغة.

ولإقامة شبكة الدِّلالات المعْجَمية بشكل مباشر حسب التصور المطروح عالِيَه فإنه يتعين توفير فريق لغوي يدرس كُلَّ مفردة ويبحث احتمال اتصالها ببقية المفردات وعددها = م – ١ عبر إحدى العلاقات الدلالية وعددها = ع، وعلى ذلك يُمْكِن حسابُ عدد الحالات التي يجب على الفريق فحصُها ونرمز له بالحرف ن$_{١}$ وفقَ الصيغة:

ن$_{١}$ = م × (م–١) × ع ≈ م^{٢} × ع

$$N_1 = M \cdot (M - 1) \cdot L \approx M^2 \cdot L$$

حيث أن م عادةً ما تكون عددًا كبيرًا جدًّا

وهذه الصيغة تُعبِّر عن تكلفة فادحة يكاد يستحيل توفيرها واقعيًّا، ولتوضيح ذلك نفترض على سبيل المثال عملية بناء شبكة دِلالات مُعْجَمية بهذه الطريقة المباشرة كي تغطي مئة ألف مفردة يمكن أن تصل بين بعضها البعض عشرون علاقة دِلالية، فتكون التكلفة إذن حوالي مئتي مليار حالة فحص (أي اثنان وأمامها أحد عشر صفرًا)، وبافتراض قدرة اللغوي على دراسة ألف حالة خلال الساعة فإن عدد ساعات العمل المطلوبة يبلغ مئتَيْ مليون ساعة عمل وهو ما يجاوز مئة ألف سنة عمل من فرد واحد!

ولتجاوز هذه التكلفة الباهظة يجري استدعاء مفهوم دلالي وسيط يسمى في بعض الأحيان «فئة الترادف» أو «الحقل الدِّلالي» أو «المَعْنَى الكُلِّيّ» وهو ببساطة معنى أساسي عامٌّ في اللغة مثل «الحُبّ» ، «الكُرْه» ، «العَمَل» ، «النجاح» ، «الفشل» ، «الصواب» ... إلخ، وتندرج تحت كل معنى كلي من هذه المعاني العديدُ من الكلمات المترادفة؛ وكمثال على ذلك نأخذ معنى «التأليف» (Authoring) الذي يمكن أن تندرج تحته كلماتٌ مثل {«كِتاب» ، «كِتابة» ، «مُؤلَّف»، «سِفْر» ، «زَبُور»، «مخطوط»، «مُصَنَّف»، «بَحْث» ، «أُطْروحة» ، «رِسالة» ، ...}[1]، مع ضرورة التنبيه إلى أن نفس الكلمة قد تندرج تحت أكثر من معنًى كليٍّ واحدٍ وذلك مثل كلمة «عَيْن» التي قد تأتي بمعنى «عُضْو الإبصار» أو «النَّبْع» أو «الجاسوس» أو «العَقَار» ... إلخ. ونفترض أن المعاني الكلية التي نعتمدها في اللغة تشكل فئة مغلقة ذات حجم = ح معنًى كليًّا.

وبجانب استدعاء هذا المفهوم الدِّلالي الوسيط نُنْشئ افتراضًا دِلاليًّا تقريبيًّا هامًّا مفاده أن العلاقةَ الدِّلاليةَ بين أيّ كلمةٍ مندرجةٍ تحت أي معنى كليٍّ «أ» وأيّ كلمةٍ مندرجةٍ تحت أي معنى كليٍّ «ب» هي نفسُها العلاقةُ الدِّلاليةُ بين المعنى الكلي «أ» والمعنى الكلي «ب».

---

١ – لا يخفى على فطنة القارئ العربي أنه لا تكاد توجد كلمتان عربيتان متساويتان دِلاليًّا؛ فعلى سبيل المثال عند مقارنة كلمة «كُرْه» بكَلِمة «مَقْت» بدقة يتضح أن درجة الحِدَّة في الصفة الثانية أشدُّ من الأولى – قال تعالى ﴿يا أيها الذين آمنوا لِمَ تقولون ما لا تفعلون، كَبُرَ مَقْتًا عند الله أن تقولوا ما لا تفعلون﴾ الصَّفّ ٢-٣، وعند فَحْص دِلالة «جَلَسَ» بالمقارنة مع «قَعَدَ» نجد أن فِعْل الجلوس لا يأتيه إلا الشخص القائم/ الواقف بينما لا يلزم ذلك في فعل القُعود – قال الشاعر أبو نُوَاس (قُلْ لِمَنْ يبكي على رَبْعٍ دَرَس ::: قائمًا ما ضَرَّ لو كان جَلَس)، ... إلخ. وعلى ذلك فإن أغلب اللغويين يرون أن الترادف بين كلمتين يعني عمليًّا شدة التقارب الدِّلالي بينهما وليس التساويَ التام.

ولإقامة شبكة الدِّلالات المعْجَمية «التقريبية» عبر هذا المفهوم الدلالي الوسيط فإن علينا توفير فريق لغوي يَدْرُسُ كُلَّ مفردة من بين فئة المفردات (وحجمها م مفردةً) ويبحث وقوعَها تحت كل معني كلي من فئة المعاني الكلية (وحجمها ح معنًى كليًّا)، إضافةً إلى دراسة احتمال اتصال كل معنى كلي بباقي المعاني الكلية وعددها = ح - ١ عبر إحدى العلاقات الدلالية وعددها = ع، وعلى ذلك يُمْكِن حسابُ عدد الحالات التي يجب على الفريق فحصُها ونرمز له بالحرف ن٢ وفقَ الصيغة:

$$ن_٢ = م \times ح + ح \times (ح - ١) \times ع \approx م \times ح + ح^٢ \times ع$$

$$N_2 = M \cdot S + S \cdot (S - 1) \cdot L \approx M \cdot S + S^2 \cdot L$$

وهذه الصيغة تُعبِّر عن تكلفة أقل بكثير من الأولى بافتراض نفس عدد المفردات ونفس عدد العلاقات الدِّلالية؛ ومع افتراض عددٍ من المعاني الكلية يساوي ألفًا على سبيل المثال تكون التكلفة إذن حوالي مئة وعشرين مليون حالة فحص وهو ما يكافئ ثلاثةً وستين سنة عمل من فرد واحد[1] وذلك أقل من التكلفة الأولى بنسبة تقارب ألف وستمئة وستة وستين مرةً.

ومن الجدير بالذِّكْر أن الوعاء اللغوي الذي تندرج فيه مفردات اللغة في مجموعاتٍ تحت معانٍ كلية يُشَكِّل كِيانًا مُعْجَميًّا يُعْرَف تقليديًّا باسم «مَكْنَز» (Thesaurus)، ومن بين المكانز العربية المعتبَرة تلك المراجع المشار إليها فى نهاية هذا الكتاب بأرقام [٣، ٤، ٦، ٧، ١١، ١٢] وتُعتبَر هذه المكانزُ وأمثالُها نقطةَ انطلاق لا غِنًى عنها لبناء أية شبكةٍ دِلاليةٍ مُعْجَميةٍ عربيةٍ يُعتَدُّ بها.

أما العلاقات الدلالية فمنها علاقتان حيويتان لا غِنًى عنهما لبناء أية شبكة دِلالية مُعْجَمية؛ الأولى علاقةُ «الترادف» وهي التي يرتكز عليها استغلال المفاهيم الدلالية الوسيطة كما سَبَقَ عَرْضُه عالِيَهُ، والثانية علاقة «الاشتمالية» وهي علاقة العام بالخاص؛ وتتمايز إلى علاقة «نَوْع من is-kind-of» مثل علاقة «السرير كنوع من الأثاث» وعلاقة «فَرْد من is-member-of» مثل علاقة «السنجاب كفرد من الثَّدْييات القوارض»

---

١ - أو سنة عمل واحدة من فريق به ثلاثة وستين فردًا يقسم بينهم العمل بالتساوي.

وتكوِّن علاقة الاشتمالية «التصنيف الهرمي» (Taxonomy) لشبكة الكلمات[1] مع ملاحظة أن الأفراد لا تكون إلا في آخر مستوى من هذا التصنيف.

ومن العلاقات الدِّلالية علاقات أساسية شهيرة لا يمكن أن تغيب عن أية شبكة دلالية معجمية:

أ. علاقة «التضاد» ؛ مثل علاقة معنى «الكَسَل» كمضاد لمعنى «النَّشاط».

ب. علاقة «السببية» ؛ مثل علاقة معنى «الأكْل» كسبب لتحقيق معنى «الشبع».

ج. علاقة «الشرطية» ؛ مثل علاقة معنى «الزواج» كشرط لوقوع معنى «الطلاق».

د. علاقة «الحالِّيَّة الزمانية»؛ مثل علاقة معنى «السَّهَر» الذي يحل في زمن «الليل».

هـ. علاقة «الحالِّيَّة المكانية» ؛ مثل علاقة «السَّمَك» الذي يحل في «الماء».

و. علاقة «الجزئية» ؛ مثل علاقة «الأنف» كجزء من «الوجه».

وهناك أيضًا العديد من العلاقات الدلالية التي تعد اشتقاقات تفصيلية لما سبق من علاقات دلالية أساسية، فضلاً عن أن بعض الشبكات الدلالية المعجمية تتضمن أيضًا معكوس العلاقات الدلالية المذكورة عاليه؛ ومن ذلك على سبيل المثال «المَحَلِّيَّة الزمانية» كمعكوس لعلاقة «الحالِّيَّة الزمانية» وذلك مثل علاقة «الليل» كمَحَلٍّ «للسهر»، كما أن علاقة «الكُلِّيَّة» معكوس لعلاقة «الجزئية» وذلك مثل علاقة «الوجه» «بالأنف»،... إلخ.

ومن الواضح تبعًا لصيغة حساب التكلفة ن٢ التي اشتققناها عالِيَهُ أن تكلفة بناء الشبكة الدلالية المعجمية تتزايد باطراد متسارع مع زيادة عدد العلاقات الدلالية ح بين مفرداتها.

> مما تقدَّمَ في هذا القسم يتبين لنا أن بناء شبكةٍ دِلاليةٍ مُعْجَميةٍ ذاتِ تغطيةٍ واسعةٍ لمفردات اللغة وللعلاقات الدِّلالية بينها يتطلب عملاً غزيرًا يقوم به فريقٌ متجانسٌ منظَّمٌ من اللغويين المتمرسين في علوم الدِّلالة والصرف والمَكَانِز والمَعاجِم.

١- وهو ما يناظر تصنيف الكائنات في الأنطولوجيا الذي ذكرناه في القسم الأول من هذا الفصل، والذي نعود إليه في القسم الرابع من هذا الفصل لتأصيل الشبكات الدلالية المعجمية كأنطولوجيا لغوية جزئية.

٣- المعالَجَة الدِّلالية المعْجَمية في اللغة العربية

يمكِن تصنيفُ المقارَباتِ المتَّبَعة في بناء الشبكات الدلالية المعجمية إلى الطريقتين الآتيتين اللتين تؤديان إلى نفس الغاية مع بعض الاختلافات التفصيلية:

أ. تُسمَّى المقارَبةُ الأولى اصطلاحًا «شبكة الكلمات» (WORD NET) [٢٧] حيث تُجسِّد فيها «فئةُ المترادفات» (Syn-Set)[1] المفهومَ الدلاليَّ الوسيطَ المشارَ إليه في القسم السابق فيُمْكِنُ مثلاً تشكيلُ فئةِ ترادُفٍ مما يلي من كلماتٍ مترادفةٍ (متقاربةٍ دِلاليًّا) {«مُعلِّم» ، «مُدرِّس»، «أستاذ»، «مؤدِّب»، «مُربِّي»، «شَيْخ»، «مُحاضِر» ... إلخ}، وكما ورد في القسم السابق فإن المفرَدة ذات الدِّلالات المتعددة قد تنتمي لأكثر من فئة مترادفات واحدة، ويتم الحصولُ على فئات المترادفات هذه عَبْرَ استقراء حصيلة المفردات وربما مع الاستعانة ببعض المَكَانِز الموسَّعة كنقطة انطلاق، ونتيجةَ استقراء المفرَدات في الاتجاه من أسفل (المفردات) إلى أعلى (فئات المترادِفات) فإن هناك ميلاً إلى كِبَرِ عدد فئات المترادفات ح. أما العلاقات الدِّلالية من مثل الأنواع التي وردت في القسم السابق فهي تصل بين فئات المترادفات نِيابةً عما تتضمنه من مفردات وهذا على وجه التقريب، وهو كما بَيَّنَّا ما يجعل تكلفة بناء شبكة الكلمات ممكِنةً من الناحية العملية.

وقد بدأ تطوير «شبكة كلمات اللغة الإنجليزية» (Princeton Word Net) في جامعة برينستون عام ١٩٨٥م [٣١] كأول شبكة دِلالات مُعجَمية محوسَبة تتَّبِع أسلوب «شبكة الكلمات» ، ثم توالى منذ ذلك الحين تطويرُ شبكات الكلمات لتغطي العديد من لغات العالم تحت مظلة «شبكة الكلمات العالمية» (Global WordNet) [٣٠] التي تضم الآن عددًا واسعًا من اللغات الحية (والأوروبية منها بصفة خاصة) بما فيها اللغة العربية. ومن الجدير بالذكر أن منظومة شبكة الكلمات العالمية تعتمد المواصفات والمعايير المتَّبَعة في كل من مشروعَيْ «شبكة الكلمات الإنجليزية من برينستون» ، و«شبكة الكلمات الأوروبية» (Euro Word Net) [٢٧].

١- الاسم الاصطلاحي Syn-Set هو مختصَر شائع للاسم الكامل Synonyms-Set.

ب. أما المقارَبة الثانية فهي «الرَّبْطُ الدِّلالي المعجَمي عَبْرَ الحقول الدلالية» ، حيث يُجَسِّدُ فيها «الحقلُ الدِّلاليُّ» (Semantic Field) [٢٤] المفهومَ الدِّلاليَّ الوسيطَ، ويمثِّلُ كُلُّ حقل دِلالي أحدَ المعاني الكلية الأساسية القائمة بذاتها مثل «الحُبّ»، «العَمَل»، «المنافَسة»، «المال» ، «السِّلاح»، «المَسْكَن»، «الدِّراسة»، «الطعام» ... إلخ، ويجري إدراج كل مفردة من مفردات اللغة تحت الحقل الأكثر تعبيرًا عن دِلالتها، ويمكن بالطبع أن تقع نفسُ المفردة ذات المعاني المتعددة تحت أكثر من حقل دلالي واحد، ومن أجل إتمام تسكين كل المفردات تحت الحقول الدِّلالية فإنه تجري الاستعانة بالمَكَانِز الموسَّعة. ونظرًا لأن إنشاء الشبكة الدِّلالية عبر هذه الطريقة يتمدد من الأعلى (أي الحقول الدِّلالية) إلى الأسفل (المفردات) فإن هناك ميلاً إلى تحجيم عدد المفاهيم الدلالية الوسيطة ح. أما العلاقات الدِّلالية فهي تصل بين الحقول الدِّلالية نِيابةً عما يقع تحتها من مفردات وهذا أيضًا على وجه التقريب، وهو ما يجعل التكلفة الاقتصادية لهذه المقارَبة أكثر جاذبية.

وبالنظر إلى ما جرى إنجازُه بالفعل من شبكاتِ دلالاتٍ معجَميةٍ لِلُّغة العربية، فإننا لا نجد حتى تاريخ كتابة هذه السطور(١) سوى القليل من النماذج التي يمكن الاعتداد بها ونذكر منها اثنين هما الأكثر اكتمالاً من حيث تغطية المفَردات العربية وثراء العلاقات الدِّلالية بينها:

٣,١- **أما النموذج الأول** فهو «شبكة الكلمات العربية» (Arabic Word Net) [١٦] وقد جرى تنفيذه حسب الأسلوب الأول المذكور عالِيَهُ؛ «شبكة الكلمات». وقد بدأ العمل في شبكة الكلمات العربية عام ألفين وخمسة ميلادية كأحد مشروعات «شبكة الكلمات العالمية» (Global Word Net)، وقد جرى ابتداءً إنشاءُ المحتوى المعجَمي لشبكة الكلمات العربية نقلاً (عبر الترجمة) عن محتوى الإصدارة الثانية لشبكة برينستون للكلمات الإنجليزية [١٧]. ويضم محتوى شبكة الكلمات العربية حوالَيْ تسعة آلاف فئة ترادف وحوالَيْ عشرين ألف مفرَدة (أو تعبيرًا مركَّبًا) وحوالَيْ مئة ألف من الأزواج المرتَّبة (مفرَدة، فئة ترادف)، كما أُجْرِيَتْ بعضُ التحسينات على هذا المحتوى المعجمي لتوسعة تغطيته وخصوصًا عبر توظيف أدوات معالجة لغوية لتمكين التعامل بمرونة

١- أي في غضون الفترة ٢٠١١م/ ٢٠١٢م حيث جرى تأليف الطبعة الأولى من هذا الكتاب.

مع البنية الصرفية للكلمات العربية[1]. وربما كانت أهم أوجه النقد التي تُوجَّه لشبكة الكلمات العربية هي أن محتواها لم يُفَصَّلْ من الأساس على مقاس اللغة العربية بل جرت ترجمتُه عن اللغة الإنجليزية، وكذلك ما واجه تطويرَها من تعثر التمويل نهايةَ عام ٢٠٠٧م/ بدايةَ عام ٢٠٠٨م. ورغم أن فريق العمل القائم على تطوير شبكة الكلمات العربية ما زال يعتبرها على طريق طويل من التطوير والتحسين[2] فإن هذا العمل يبقى ذا قيمة معتبَرة خاصةً وأنه متاح بشكل مفتوح المصدر للجميع.

٣,٢- وأما النموذج الآخَر فهو «قواعد بيانات الدِّلالات المعجَمية العربية عبر الحقول الدِّلالية» (Arabic Lexical Semantics Database via Semantic Fields) [١٥] والذي جرى تنفيذُه حسبَ الأسلوب الثاني المذكور عالِيَهُ «الرَّبْطُ الدِّلالي المعجَمي عَبْرَ الحقول الدِّلالية» بواسطة «الشركةُ الهندسية لتطوير نُظُم الحاسبات (آرْ-دِي-آيْ)» (RDI; www.RDI-eg.com) التي عكفت على بناء وتطوير هذه الشبكة الدِّلالية المعجمية العربية بين منتصف عام ٢٠٠٥م حتى بداية عام ٢٠١١م بدعمٍ وتمويلٍ جزئيٍّ من «مركز تميز التنقيب المعلوماتي والنمذجة الحاسوبية http://dmcm.cu.edu.eg» بمصر بين بداية عام ٢٠٠٦م إلى بداية عام ٢٠١٠م [٣٤]. وقد مر العمل عبر هذه السنوات بثلاث مراحل، حيث امتد الإنجازُ مع إتمام المرحلة الثالثة في بداية عام ٢٠١١م ليشمل نحو مئة ألف مفردة عربية صريحة تندرج تحت مستويين من الحقول الدلالية؛ مستوًى أساسيٍّ عام به نحو ألفين من المعاني والمفاهيم الكلية، ومستوًى ثانويٍّ يتشعب إلى حوالَيْ ثمانية عشر ألفًا وأربعمئة حقلٍ دِلاليّ أكثر تفصيلاً لتصنيف المعانيَ الكليةَ وفقَ الصيغ الصرفية[3] ووفقَ بعض التصنيفات الدِّلالية[4].

١- توظِّف شبكةُ الكلمات العربية الآن المحلِّلَ الصرفي العربي «تِيمْ باكْوُولْتَر Tim Buckwalter» http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2002L49 وذلك لربط حصيلة واسعة من الكلمات العربية الخام التي قد يُطْلَب الاستعلام عنها دِلاليًّا بالحصيلة المحدودة الواردة على نحوٍ صريحٍ في الشبكة (راجع فصل «التحليل الصرفي الآلي للمفردات العربية» من هذا الكتاب).

٢- انظر http://www.vossen.info/docs/2008/ArabicWordNet-GWN2008-final.pdf.

٣- هذه الأصناف الصرفية الأربعة عشر هي {اسم آلة، اسم تفضيل، اسم ذات، اسم زمان، اسم فاعل، اسم معنى، اسم مفعول، اسم مكان، اسم منسوب، صفة مُشبَّهة، صيغة مُبالَغة، فِعْل، كلمة وظيفية، مَصْدَر}.

٤- هذه التصنيفات الدِّلالية هي {حَدَث، مَعْنَى، فاعل، مفعول، صفة مشبَّهة، تفضيل، ذات، مكان، زمان، كلمة وظيفية}.

وترتبط هذه الحقولُ الدِّلالية بعضُها ببعضٍ عبرَ عشرين نوعًا مما أشرنا إليه من علاقات دِلالية في نهاية القسم السابق سواء أكانت علاقات أساسية أم مشتقة أم معكوسات هذه وتلك، وتتحكم التصنيفاتُ الصرفيةُ والدِّلاليةُ للحقول الدِّلالية في المستوى الثانوي فيما ينطبق وما لا ينطبق بينها من علاقات دِلالية؛ فعلى سبيل المثال لا تقوم علاقةُ الحالِّيَّة الزمانية إلا بين حدثٍ وزَمَن، ونتيجةً لذلك فإن قيمة ح الفَعَّالة المستخدَمة في حساب تكلفة إنشاء الشبكة الدِّلالية هي ألفان وليست ثمانية عشر ألف وأربعمئة. وفي محصلة ذلك كله فإن لدينا في نسيج الشبكة أكثر من مئتين وثلاثة وتسعين ألفَ وصلة دِلالية تربط الحقول الدِّلالية بعضَها ببعض.

ومن الجدير بالذكر أن المئة ألف مفردة المندرجة في قاعدة بيانات الربط الدلالي المعجمي هذه ليست مجرد كلمات عربية نهائية وإنما هي مُركَّبات صرفية مرنة يمثِّلُ كلٌّ منها عددًا كبيرًا من الكلمات النهائية المكوَّنة من تباديل هذه المركَّبات مع اللواصق؛ مثل «مَكْتَب» التي تمثِّل أيضًا {المكتب، لمكتبه، فالمكتبان، مكتبنا، بمكتبي، ...}، وذلك تحتَ قُيودٍ تضمنُ بقاءَها ضمن نفس دِلالتها وتستخدم آرْ-دِي-آيْ لهذا الغرض أدواتِها للتحليل الصرفي والعنونة النحوية [١٤]، وعلى ذلك فإن هذه المئة ألف من المفردات تغطي فعليًّا عشرات الملايين من الكلمات العربية النهائية مما يؤشر على مراعاة طبيعة بِنْية اللغة العربية في تصميم هذه الشبكة الدِّلالية المعجَمية. وقد كانت نُوَاة إنشاء محتوى هذه الشبكة الدِّلالية هي «المُكْنَز الكبير» [٤] وهو عمل قَيِّم أنجزه فريق كبير من الباحثين تحت قيادة عالم اللغة الشهير الراحل «أ.د. أحمد مختار عمر» لرصد الحقول الدِّلالية الأساسية في اللغة وإدراج المفردات العربية تحتها، ثم أخِذَتْ بعين الاعتبار عدةُ مصادر أخرى فيما بعد عند تنقيح محتوى الشبكة الدِّلالية [٣، ٦، ٧، ٨، ٩، ١٠، ١١، ١٢] فضلاً عن الاستفادة من «شبكة الكلمات العربية» (Arabic Word Net) السالف ذكرُها.

٤- شبكات الدِّلالات المُعْجَمية كإطارٍ أُنْطُولوجيٍّ جزئي

يُمكِنُنا الآن تأصيلُ الشبكات الدِّلالية المعْجَمية - سواء لعبت فئات الترادف أم الحقول الدِّلالية دَوْرَ المفاهيم الدِّلالية الوسيطة فيها - كأنطولوجيات جزئية وفق العرض الذي قدمناه في القسم الأول من هذا الفصل، وذلك بالترتيب على النحو التالي:

أ. كائنات الأنطولوجيا وأفرادُها تناظرُها المفرداتُ في الشبكة الدِّلالية.

ب. الفئاتُ/ الأنواعُ في الأنطولوجيا تناظرُها في الشبكة الدِّلالية المفاهيمُ الدلاليةُ الوسيطة - سواء أكانت فئاتِ ترادفٍ أو حقولاً دِلالية - كما أن التصنيف الهرمي لهذه الفئات في الأنطولوجيا تمثلُه في الشبكة الدِّلالية علاقةُ الاشتمال.

ج. العلاقات بين كائنات الأنطولوجيا تناظرُها العلاقاتُ الدِّلاليةُ بين المفاهيم الدِّلالية الوسيطة - سواء أكانت فئاتِ ترادفٍ أو حقولاً دِلالية - وبالتبعية بين ما يندرج تحتها من مفرداتٍ وهو التقريبُ الدِّلاليُّ العمليُّ الذي ذكرناه.

د. خصائصُ الفئاتِ/ الأنواعِ والكائناتِ تحتَها في الأنطولوجيا تناظرُها الخصائصُ الصرفيةُ والدِّلاليةُ (المذكورة في الملحوظتين الهامشيتين رقم ١١ و ١٢ عالِيَهُ) التي تؤدي لتفصيل الحقول الدلالية الأساسية إلى حقول دلالية ثانوية.

ومن الملاحَظ أن عناصرَ الأنطولوجيا الحاضرةَ في الشبكات الدِّلالية المعجمية هي تلك المكوِّنات الثابتة (الإِسْتاتِيكِيَّة) بينما تغيبُ بقيةُ عناصر الأنطولوجيا من مكوِّناتٍ أنطولوجيةٍ متحركة (دِينامِيكِيَّة) إضافةً إلى تلك التي تمثل حقائقَ العالم الخارجي العامةَ المعلومةَ خارج إطار النص لكنها حاكمةٌ لفهمنا الدِّلالي له، وما لم تُتِمّ الأبحاث والتقنياتُ الحالِّيَّةُ إنجازَ هذه المكونات الأنطولوجية الديناميكية فإن الطريق لا تزالُ طويلةً نحو تجاوز المستوى المعجَمي في التحليل الدِّلالي لِلُّغة الحية.

٥- الالتباس الدِّلالي والعمل على إزالته

يُنْتِجُ التحليلُ الدِّلاليُّ المعْجَميُّ لكل كلمة - بصفةٍ عامةٍ - عدةَ أجوبةٍ مُمْكِنةٍ وهو ما يُعْرَفُ باسم «الالتباس الدِّلالي المعجَمي» (Lexical Semantic Ambiguity)، ويرجع هذا الالتباس إلى تعدُّد المعاني الكلية للمفرَدة الواحدة مما يؤدي إلى تعدُّد الحقول الدِّلالية التي تندرجُ تحتها (أو فئات الترادف التي تنتمي إليها) هذه المفرَدةُ؛ فعلى سبيل المثال تتعدَّدُ المعاني الكليةُ للكلمات العربية «عَيْن»، «قَصْر»، «هِجاء»، «مَصْرِف»، «عَصْر» على النحو التالي بالترتيب {عُضْو الإبصار، نَبْع، جاسوس، عَقَار، ذات الشَّيْء، ...}، {مَنْع، مَسْكَنٌ فاخِرٌ متَّسِع، ...}، {الذَّمّ، تكوين الكلمات من الحروف، ...}، {مؤسَّسة مالية،

إنفاق المال، مَهْرَب، ...}، {حقبة زمنية، أحد أوقات اليوم، ...}. وتُعَدُّ مسألةُ «فك الالتباس الدِّلالي للكلمات» (Word Sense Disambiguation) إحدى المسائل العتيدة العنيدة التي لم يتم حَلُّها بَعْدُ بشكلٍ مُرْضٍ سواء في اللغة العربية أم في سواها. فبينما أثمر تطبيقُ أساليب التعلُّم الحاسوبيِّ لفك الالتباس الصرفي - عَبْرَ ترجيح التحليل ذي الاحتمال الرياضي الأعلى ضمن سياق التحليلات الصرفية الممكنة للنص المحيط به(١) - عن معدَّلات خطأ منخفضة إلى درجة مقبولة(٢) تكفي لسلاسة أداء التطبيقات المرتكِزة على هذا النوع من التحليل، فإن هامش الخطأ الناتج عن استخدام نفس الأساليب لفك الالتباس الدلالي للكلمات أوسعُ كثيرًا من ذلك حيث لم يقل في أمثل الأحوال مع اللغة الإنجليزية عن خمسة وعشرين في المئة وقد يرتفع إلى ما فوق الثلاثين في المئة مع اللغة العربية(٣) وهو ما يعرقل أداءَ التطبيقات المرتكِزة على التحليل الدلالي المعْجَمي - أو بعضًا منها على الأقل - مما سنأتي على ذكره في القسم التالي من هذا الفصل [١٣].

ويُمْكِنُ تفسيرُ اتساع هامش الخطأ في فَكِّ الالتباس الدِّلالي للكلمات عَبْرَ أساليب التعلُّم الحاسوبي مقارنةً بما تُحَقِّقُه نفسُ الأساليب مع مسألة فَكِّ الالتباس الصرفي بحقيقة أن السياقَ الاحتماليَّ الدِّلاليَّ في النص أكثرُ اتساعًا في المتوسط من نظيره الصرفي، فلا ترتبطُ دِلالةُ الكلمة فقط بالكلمات القليلة المجاورة لها ولكنها كثيرًا ما تتأثر أيضًا بكامل الفقرة(٤) التي تَرِدُ فيها (بل ربما تتأثر كذلك بما يتجاوز فقرتَها مما هو داخل وما

١- وذلك كما جرى استعراضُه في نهاية القسم الثامن من فصل « التحليل الصرفي الآلي لمفردات اللغة العربية» في هذا الكتاب.

٢- حيث يُمْكِن للمحلِّلات الصرفية العربية رفيعة الأداء أن تصل بهذا المُعدَّل للخطأ إلى ما دون الخمسة في المئة عند التدريب والتقويم على كامل اتساع الفضاء اللغوي، بينما يُمْكِن أن ينخفض إلى نحو الثلاثة في المئة عند التعامل مع حيز ضيق من الفضاء اللغوي، وذلك أيضًا كما جاء ذِكْرُه في نهاية القسم العاشر من فصل «التحليل الصرفي الآلي لمفردات اللغة العربية» في هذا الكتاب.

٣- علينا الانتباه إلى أن هامش الخطأ في فك الالتباس الصرفي العربي داخلٌ أيضًا في هامش خطأ فك الالتباس الدِّلالي للكلمات العربية؛ فنظرًا لِقلة عدد المفرَدات المذكورة بشكل صريح في أية شبكة للدِّلالات المُعجَمية (من أجل الحفاظ على تكلفة معقولة لبنائها) مقارنةً بحجم الحصيلة الكلية للمفرَدات العربية، فإن الكلمة العربية المطلوب تحليلها دِلاليًّا لا بُدَّ لها أن تمر أولاً على محلِّل صرفي يَرُدُّها إلى إحدى مشتقاتها المذكورة صراحةً في قاعدة بيانات شبكة الدِّلالات المُعجَمية (كما سبق ذِكْرُه في القسم الثالث من هذا الفصل).

٤- هذه الكلمة «فقرة» في هذا الموضع مثالٌ بَيِّنٌ على ذلك؛ حيث لا نستطيع تقرير ما إذا كان معناها المقصود هو «جزء من النص» أو «جزء من برنامج» أو «حلقة في العمود العظمي للحيوان الفقاري» ... إلخ إلا إذا وسعنا إطار السياق المحيط بها لكي يشمل قرائن من كلمات تسبقها بمسافةٍ بعيدةٍ مثل «نَصّ» و «بالكلمات» ويشمل كلمات تلحقها بمسافة بعيدة مثل «كَلِمة».

هو خارج النص)؛ فإذا كان يُمْكِن تحقيقُ هامِشِ خطأٍ صغيرٍ في عملية فك الالتباس الصرفي بدراسة سياق احتمالي ضيق لكل كلمة يمتد لكلمتين أو ثلاثًا مما يسبقها وأخرى مثلها مما يلحقها، فإن تحقيقَ هامشِ خطأٍ مقارِبٍ لذلك في عملية فك الالتباس الدِّلالي ربما يحتاجُ لدراسة نحو عشرة أو عشرين كلمة مما يسبقها وأخرى مثلها مما يلحقها، وهو ما يتجاوز بكثير القدرات الراهنة للمعالجات الحاسوبية[1].

وقد تستطيعُ بعضُ التطبيقات المرتكزة على التحليل الدِّلالي المعجَمي أن تتجاوز ذلك المعدل المرتفع للخطأ في فك الالتباس الدلالي آليًّا عبر استغلال بعض القيود الإضافية التي تنشأ عند استدعاء الدِّلالات المعجَمية في مثل تلك التطبيقات، ومنها على سبيل المثال:

٥ ,١- الاستفادةُ من ضِيق النِّطاق اللغوي للتطبيق عَبْرَ إنشاء شبكة دِلالية مُعْجَمية مكرَّسة لذات النِّطاق اللغوي الضَيِّق؛ فيُمكِن مثلاً تنفيذُ ذلك في تطبيقٍ لاسترجاع المعلومات حول «الآفات التي تصيب الحاصلات الزراعية الصيفية في صعيد مصر» فالبعدُ الجغرافيُّ مقيَّد بصعيد مصر والزَّمَنيُّ مقيَّد بالزمن المعاصِر والموضوعيُّ مقيَّد بآفات الحاصلات الزراعية الصيفية والأسلوبيُّ غالبًا سرديٌّ تقريريٌّ، ولذلك فإن كلماتٍ مثل «فِطْر» و «الجِذْر» و «حرارة» سوف تقتصر معانيها - على الترتيب - في الأنطولوجيا التي يرتكزُ عليها هذا التطبيقُ على «نوع من النباتات اللا خَضَرية التي تتكاثر لا جِنْسيًّا» (مع تراجُع معاني {«تناول طعام بعد امتناعٍ عنه»، ...}) ثم «القسم الأسفل من النبات تحت الأرض» (مع تراجُع معاني {«الأصل»، «السبب»، «القاعدة»،...}) ثم «السخونة» (مع تراجُع معاني {«شدة العاطفة»، «الحماس والشغف»، «علامة على استعداد الآلة للعمل»، ... }).

٥ ,٢- الاستفادةُ من عدم الحاجَة أحيانًا إلى الحصول على المعاني الأكثر تفصيلاً للكلمات والاكتفاءُ بالمعاني الأعم لها؛ فقد يَصْلُحُ ذلك مثلاً في تطبيقٍ لتبويب المقالات آليًّا تحت عدد محدود من الأقسام العامة في إحدى المواقع الصحفية الإلكترونية. وفي هذه

١- تتضاعف القدراتُ الحاسوبيةُ المطلوبةُ لمعالجة سياق احتمالي مع زيادة اتساع هذا السياق وفقَ نمطٍ أُسِّيٍّ وليس خطيًّا، ولذلك فإن الفارق مهولٌ بين القدرات الحاسوبية المطلوبة لدراسة سياقٍ احتماليٍّ باتساع خَمْسِ كلمات وبين سياق احتمالي باتساع نحو أربعين كلمة (أو حتى نحو عشرين كلمة).

الحالة يُكْتَفَى بتعيين الحقول الدِّلالية للكلمات في المستويات العُلْيَا (العامة) من شجرة التصنيف الدِّلالي والتخفُّف من عِبْء البحث في غابة المستويات الدنيا (التفصيلية) من هذا التصنيف مما يؤدي في المحصلة إلى التخفيف من درجة الالتباس الدِّلالي ومن ثَمَّ تقليل فُرَص الخطأ في عملية إزالته؛ ويوضِّح الجدول رقم (٥-١) أدناه بعض الأمثلة على ذلك:

الكَلِمة	حقولها الدِّلالية التفصيلية	حقولها الدِّلالية العامَّة
مَسْجِد	{محَلّ الصلاة، مكان أو مبنى مخصَّص للصلاة}	{مكان الصلاة}
نَقْد	{دِراسة العيوب والحسنات، المال، التمويل، دَفْع المال، أوراق مخصوصة للبيع والشراء}	{دِراسة العيوب والحسنات، المال}
مَطْعَم	{ما يؤكل من طعام، مكان مخصَّص لتناول الطعام}	{تناول الطعام}
يَخْشَى	{الرهبة، الخوف، التوجُّس والقلق}	{الخوف}

الجدول ٥-١ : أمثلة على الحقول الدِّلالية التفصيلية والعامة لبعض الكلمات العربية.

٥,٣- الاستفادةُ في تطبيقاتٍ مثل «استرجاع المعلومات» (Information Retrieval) من الجِسْم النَّصِّيّ محل البحث كمُرَشِّح («فِلْتَر Filter») للدِّلالات المعجَمية لكلماتِ «طلبِ البحث» (Query) وذلك قبل إجراء عملية فك الالتباس الدِّلالي لها؛ حيث يُمْكِنُ استبعادُ بعضِ الحقولِ الدِّلالية الممكِنة لهذه الكلمات في شبكة الدِّلالات المعْجَمية مما لا يتوافرُ له نظيرٌ بشكل مترابط في ذلك الجسم النصي، مما يخفف من درجة الالتباس الدِّلالي وهو ما يؤدِّي بدوره إلى رَفْعِ دِقَّةِ عملية فك الالتباس الدِّلالي التي تلي ذلك الاستبعاد.

ويبقى بطبيعة الحال جانبٌ كبيرٌ من التطبيقات الأخرى الهامة (مما سنأتي على ذِكْرِهِ في القسم التالي من هذا الفصل) تحتاجُ للتعامل مع الفضاء اللغوي باتساع نطاقاته على مختلَف الأبعاد ولذلك فإن ارتفاع معدَّل الخطأ في عملية فك الالتباس الدِّلالي ينعكس على فعالية تلك التطبيقات التي لا مَفَرَّ أمامَها سوى الانتظار حتى تنضجَ الأبحاثُ الجاريةُ لمعالجة اللغة حاسوبيًّا على مستوياتٍ أعمقَ تجمع بين توسيع شبكات الدِّلالات المعجَمية بإضافة مكوِّناتٍ مناظِرةٍ لبعض العناصر الأنطولوجية الديناميكية (مما جاء في القسم الأول في هذا الفصل) إضافةً إلى تحليل التركيب النحوي العميق للنص [١٣]

وكل هذا بالطبع في إطارٍ من أساليب التعلُّم الحاسوبية، وهذا ما سنأتي على ذِكْرِهِ في القسم الثامن من هذا الفصل.

٦- تطبيقات التحليل الدِّلالي المُعْجَمي

قبلَ استعراض أبرز تطبيقات التحليلِ الدِّلاليِّ المعْجَمي نحتاجُ إلى التعرُّف على مفهوم «الترابط الدِّلالي» الذي ترتكزُ عليه العديدُ من هذه التطبيقات؛ حيث تُقَاسُ «المسافةُ الدِّلاليَّة» بين كلمتين بأقل عدد من «الوَصَلات» (Links) بين الحقلين الدِّلاليَّيْن اللَّذَيْن تنتمي إليهما كلتا الكلمتين في «التصنيف الدلالي الهرمي» (Taxonomy)، وقد يُدْخِلُ بعضُ الباحثِين أحيانًا في حساب هذه المسافة الدِّلالية أوزانًا نسبيةً لهذه الوَصَلات تزيد شيئًا ما مع الاقتراب من قِمَّة هرم التصنيف وتقل شيئًا ما مع الاتجاه نحو قاعدته، ويتناسب «الترابطُ الدِّلالي» بين كلمتين تناسبًا عكسيًّا مع المسافة الدِّلالية بينهما.

وبمعرفةِ ذلك يُمْكِنُنا إلقاءُ الضوء فيما يلي على أمثلةٍ هامةٍ لتطبيقات مُعالجَة اللغة حاسوبيًّا التي ترتكزُ على التحليلِ الدِّلاليِّ المعجَميِّ:

أ. المساعدَة في اختيار التحليل النحوي (الإعراب) السليم من بين الاحتمالات الغزيرة التي عادةً ما تتوالدُ عند محاولة إجراء التحليل النحوي حاسوبيًّا، وذلك ما تَطْرَحُه المقولةُ الشائعة «الإعرابُ فَرْعُ المعنى» التي تختصرُ الارتباطَ العضويَّ بين كلٍّ من الدِّلالة المقصودة للنص وتركيبِها النحوي الصحيح؛ فإذا دَرَسْنا على سبيل المثال النَّصَّ التالي «... وفي الوقت الذي تعاني فيه اقتصادات البلدان <u>الأفريقية</u> من مشكلات ...» للمفاضلةَ بين إعرابَيْ كلمة «الأفريقية» إما نعتًا مرفوعًا لكلمة «اقتصادات» أو نعتًا مجرورًا لكلمة «البلدان»، فقد تَرْجُحُ كفةُ الإعراب الثاني بسبب تفوُّق الارتباط الدِّلالي المعجَمي بين كلمتَي «البلدان» و «الأفريقية» (حيث قد تندرج مفرداتٌ مثل {بَلَد، بُلْدَان ...} تحت حقل دِلالي من قَبِيل «وحدة جُغْرافية أو سياسية» بينما قد تندرج مفرداتٌ مثل {أفريقيا، أفريقية ...} تحت حقل دِلالي قريب منه من قَبِيل «منطقة جغرافية») تَفَوُّقًا ملحوظًا على الارتباط الدِّلالي الأضعف بين الحقل الدِّلالي الذي تنتمي إليه كلمة «اقتصادات» وذلك الذي تنتمي إليه كلمة «الأفريقية».

ب. المساعَدة أيضًا في اختيار التحليل الصرفي الصحيح من بين الاحتمالات العديدة التي عادَةً ما تتولدُ عند محاولة إجراء التحليل الصرفي الحاسوبي لكلمةٍ ما؛ فإذا دَرَسْنا مثلاً كلمة «بطين» في النص التالي «... فيتدفق الدم في البطين الأيمن من القلب ...» فإن الحقولَ الدِّلاليةَ التي تقع تحتَها الكلماتُ الواقعةُ في سياقها النصي القريب مثل «الدَّمُ» و «القَلْب» ذاتُ ارتباطاتٍ دِلاليةٍ أقوى مع الحقل الدِّلالي الذي يقع تحتَه التحليلُ الصرفيُّ «بُطَيْن» (على صيغة التصغير فُعَيْل) من تلك الارتباطاتِ الدِّلاليةِ مع بقية التحليلات الصرفية المحتملة الأخرى مثل التحليل «بَطِين» (على صيغة المبالغة فَعِيل) ومثل التحليل «بِطِين» (كمُرَكَّب جارّ ومجرور: بِـ + ـــطِين) ... إلخ، وعلى ذلك يَرْجُحُ التحليلُ الصرفي «بُطَين» على غيره من التحليلات الصرفية القياسية المحتملة الأخرى لكلمة «بطين».

أ. وبنفس منطق التطبيقين السابقين فإن دراسة الارتباطات الدِّلالية المعجَمية بين كلمات النص الناتج عن تقنيات التعرف الآلي؛ مثل «التعرف الآلي على الكلام المنطوق» أو «القراء الآلية للنص المكتوب» ... إلخ، لها قيمةٌ ثمينةٌ في تقدير مدى صِحة مخرَجات هذه التقنيات حيث تَكْمُن الفكرةُ في أن تَهاوِيَ متوسط قيمة الارتباط الدِّلالي لإحدى الكلمات بما يجاورُها من كلمات في سياقها النصي تحت المتوسط العام لقيمة هذا الارتباط بشكل كبير قد يؤشِّر إلى خطأ في التعرف على هذه الكلمة؛ فالعينةَ النصية التالية على سبيل المثال والتي قد يُنْتِجُها نظامٌ للتعرف الآلي على الكلام المكتوب «... ولذلك تعلو قيمة القرد المبدع بإسهاماته الابتكارية في المجتمعات الحديثة ...» تحتوي على كلمة «القرد» ذات الترابط الدِّلالي الضعيف مع باقي كلمات هذه العينة النصية مما يثير الرِّيبةَ في صحة التعرف عليها ومن ثم الرجوع إلى نظام التعرف الآلي مرة أخرى للبحث عن بديل آخر ذي ترابط دلالي قوي بما حوله ربما يكون كلمة «الفرد» في المثال السابق.

ج. يُمْكِنُ أيضًا تطبيقُ نفس المنهج لمراجعة مُخْرَجات أنظمة الترجمة الآلية، وبصفةٍ عامةٍ فإن تضمينَ دراسة الارتباط الدِّلالي المعجَمي داخلَ آليات البحث والمفاضَلة الاحتمالية بين الوحدات اللغوية المُمْكِنة أثناءَ توليد النص (المتعرَّف

عليه أو المترجَم) في كُلٍّ من هذه الأنظمة يكون أكثر فعالية من استخدامها كمرحلة مراجعة نهائية بعد توليد ذلك النص، حيث يسمحُ هذا التضمينُ باستدعاء بدائل أخرى إلى أن يتحققَ تعظيمُ متوسط الترابط الدلالي بين كلمات النص المترجَم أو المتعرَّف عليه ... إلخ.

هـ. يمكن لأنظمة «التدقيق والتصحيح الإملائي» (Spell Checking and Correction) أن تستفيد كذلك من دراسة الارتباط الدِّلالي بنفس الطريقة لتعيين الكلمات المشتبَه في خطئها رغم سلامتها هجائيًّا حيث قد يتعذر اكتشاف مثل هذا النوع من الأخطاء بأية وسيلة أخرى.

و. تحسين أداء أنظمة «استرجاع المعلومات النصية» (Text Information Retrieval) بما فيها وظائف البحث النصي، وذلك عبر فهرسة الأوعية النصية مَحَل البحث بالدِّلالات المعْجَمية لكلماتها بدلاً من فهرستها بالمركبات الصرفية لهذه الكلمات من جذورٍ أو جذوع[1]، ويرجع هذا التحسن إلى أن الفهرسة بالدِّلالات المعْجَمية قد ترفع كلا المؤشِّرين اللذَيْن يقاس بهما أداءُ مثل هذه الأنظمة وهما:

- مؤشِّر الاستدعاء» (Recall)[2] حيث تعمل العلاقات الدِّلالية وخصوصًا علاقة الترادف على زيادة القدرة على استدعاء الكلمات المشابهة دِلاليًّا لكلمات طلب البحث حتى لو لم يكن بينها علاقةٌ صرفية؛ فعلى سبيل المثال إذا وردت كلمةُ «الأسد» في طلب البحث فإننا نتحصَّلُ نتيجةَ الفهرسة الدِّلالية على كلماتٍ من قَبيل {اللَّيْث، الغَضَنْفر، الهَيْثَم، ...} (إضافةً بالطبع إلى «الأسد» ومشتقاتها) مما قد يَرِدُ في الأوعية النصية محل البحث، وهذا بالتأكيد يتجاوزُ ما نحصل عليه نتيجة الفهرسة الصرفية بمُفْرَدها.

١- وذلك كما ورد في الفقرة أ من القسم السادس من فصل «التحليل الصرفي الآلي لمفردات اللغة العربية» في هذا الكتاب.

٢- يُعرف «مؤشر الاستدعاء» كميًّا بالإشارة إلى «طلب بحثٍ» ما في قاعدة بيانات على أنه خارج قسمة «عدد المعلومات المستدعاة – ضمن كامل فئة المعلومات المستدعاة – التي تعتبَر استجاباتٍ صحيحةً لطلب البحث» على «عدد كل المعلومات التي تعتبر استجابات صحيحة لطلب البحث في كامل قاعدة البيانات محل البحث».

- «مؤشِّر الدِّقَّة» (Precision)(١) حيث يجري استبعادُ الكلمات البعيدة دِلاليًّا عن كلمات طلب البحث حتى مع وجود تشابه بنائي (صرفي)؛ فعلى سبيل المثال إذا وردت كلمةُ «عَيْن» في طلب البحث بمعنى «مصدر جوفي للمياه» فإن الفهرسة الدِّلالية للأوعية النصية محل البحث تؤدي إلى استبعاد كلماتٍ متشابهة صرفيًّا من قَبيل {أعْيُن، عيون (بمعنى جواسيس)، عَيْنِيّ ...} مما قد يَرِدُ في هذه الأوعية، بينما تؤدي في الوقت نفسه إلى الحصول على كلماتٍ من قَبيل {نَبْع، جَدْوَل، عُيُون (بمعنى ينابيع)، بئر، ينابيع ...} مما قد يَرِدُ في الأوعية النصية مَحَلّ البحث.

ز. كما يجب أن نؤكِّد على أن التحليل الدِّلالي المعجَمي خطوةٌ أولى تأسيسةٌ نحو المعالجة الدِّلالية العميقة والممتدة أيًّا كانت المقارَبات المستخدمة فيها.

وفي حين أن القائمةَ السابقة قد تتسع للمزيد من تطبيقات التحليل الدِّلالي المعجَمي، فإن الملمح الواضح المشترك بينها هو احتياجُها جميعًا لدرجة مقبولة من الدقة في فك الالتباس الدِّلالي المعْجَمي وهو التحدي الذي لم تتغلب عليه أساليبُ المعالجة الحاسوبية لِلُّغةِ بشكلٍ مُرْضٍ حتى الآن.

١ - يُعرف «مؤشر الدقة» كميًّا بالإشارة إلى «طلب بحثٍ» ما في قاعدة بيانات على أنه خارج قسمة «عدد المعلومات المستدعاة - ضمن كامل فئة المعلومات المستدعاة - التي تعتبَر استجاباتٍ صحيحةً لطلب البحث» على «عدد كل المعلومات المستدعاة في فئة المعلومات المستدعاة».

٧- العَنْونة الدِّلالية المُعْجَمية للمدوَّنات النصية العربية

الشَّكل ٥-١ : لقطة من أداة عنونة النص العربي صوتيًّا وصرفيًّا ودِلاليًّا؛ «فَصيح»[١]، ويلاحَظ أن اختيار الحقل الدِّلالي الأنسب للكلمة «مَرّ» يجري بعد حسم تحليلها الصرفي.

يوجَدُ الالتباسُ الدِّلالي المعْجَمي كأمرٍ واقعٍ في أية شبكةِ دِلالاتٍ مُعجَمية، وتحتاجُ تطبيقاتُ التحليل الدِّلالي المعجَمي كما ورد في القسم السابق من هذا الفصل إلى فك هذا الالتباس، ولا غِنًى في الوقت الراهن عن الاستعانة بأساليب التعلُّم الحاسوبي لفك هذا الالتباس، وتحتاجُ مثلُ هذه الأساليب بدورها إلى موارِد لغوية كي تتدرب عليها وتُكَوِّنَ تبعًا لذلك نماذجَها الاحتمالية الرياضية، وتتكوَّنُ هذه المواردُ في حالتنا هذه من نصوص تُحَدَّدُ لكل كلمة فيها الحقلُ الدِّلالي (أو فئةُ الترادف) الأنسب لها، ويُسَمَّى تجهيزُ هذه الموارِد بعملية «العنونة الدِّلالية المعجَمية» للمدوَّنات النصية التي ينطبق عليها ما ورد في القسم التاسع من باب «التحليل الصرفي الآلي للمفردات العربية» في هذا الكتاب. وبالإضافة إلى ذلك ينبغي الانتباه إلى النقاط التالية:

١- تحتوي الورقة المشار إليها برقم [١٤] في قائمة مراجع هذا الكتاب على تفاصيل هذه الأداة التفاعلية.

- لا تجري عمليةُ العنونة الدِّلالية المعجَمية إلا على نَصٍّ جرى حسمُ التحليل الصرفي لكلماته (وتحددت أيضًا «أنواعُها الكلامية PoS tags»)، ويعرضُ الشكلُ رقم 1 أعلاه لقطةً أثناء تشغيل إحدى أدوات عنونة النصوص العربية التي تدعمُ كلَّ هذه العَنْوَنات.
- يقوم اللغويون المؤهَّلون بهذه العنونة الدِّلالية المعجَمية للنصوص، وأحد المؤهلات المطلوبة هي الإلمامُ بقدر معقول من المعرفة عن موضوع المدوَّنةِ المطلوبِ عنونتُها.
- يوجَدُ هامشُ اختلافٍ طبيعي بين اللغويين القائمين بالعنونة الدِّلالية أو بتقويم أداء أي نظام تحليلٍ دِلاليٍّ آليٍّ (على خلاف العنونات الصرفية والصوتية) ولذلك يَحْسُنُ أن يَجْرِيَ الأخذُ بالرأي الغالب لفريقٍ يتكون من عددٍ فردي من المعنونِين.

٨- التحليل الدِّلالي ما بعد المستوى المُعْجَمي

كما ألمحنا في نهاية القسم الخامس من هذا الفصل، فإن الأعمال البحثية جاريةٌ لمعالجة اللغة دِلاليًّا فيما يتجاوز المستوى المعجَمي وهي تتوجه عمومًا نحو الجمع بين توسيع شبكات الدِّلالات المعجَمية بإضافة مكوِّناتٍ مناظِرةٍ لبعض العناصر الأنطولوجية الديناميكية (مما جاء في القسم الأول في هذا الفصل) وتحليلِ التركيب النحوي العميق للنص، وكل هذا بالطبع في إطارٍ من أساليب التعلُّم الحاسوبية [١٣].

ولا تقتصرُ أهميةُ دراسة التحليل الدِّلالي العميق على السعي نحو تحقيق الأهداف الإستراتيجية الكبرى لمسيرة المعالجة الحاسوبية للغات الحية بما يمكِّن من امتلاك الآلات لناصية اللغة فهمًا وتوليدًا بمستوى يقترب (ولو جزئيًّا) من المستوى البشري[1]، بل إن هذه الأهميةَ مُلِحَّةٌ من منظورٍ عَمَلِيٍّ مَرْحَلِيٍّ كذلك من أجل رفع أداء العديد والعديد من التطبيقات الأساسية المرتكزة على التحليل الدِّلالي المعْجَمي (من مِثْل تلك المذكورة في القسم السادس من هذا الفصل) إلى مستوى يثبِّتها في مصاف التقنيات الحاسوبية التي يُعْتَمدُ عليها.

١- وإنْ تحققت مثل تلك القدراتُ اللغويةُ للآلات فإنها سوف تتفوق على البشر بكونها تتوفر على قدراتٍ هائلة لمُباشَرةِ ومُعالجَةِ أوعيةٍ شاسعةِ الاتساعِ باذخةِ الغِنى من المعلومات والمعرفة.

وضمن هذا الإطار تقع «لغةُ الشبكات الدِّلالية الحاسوبية العالمية» (Universal Networking Language) – التي تُعْرَفُ أيضًا بالاختصار (UNL) – في موقعٍ رياديٍّ بين جهود الحوسبة الدِّلالية لِلُّغةِ فيما يتجاوزُ المستوى المعْجَمي [٣٣]، وهي لغة حاسوبية مُحْكَمةٌ مصمَّمةٌ خِصِّيصًا لتمثيل البيانات الدِّلالية المستخلصة من نصوص اللغات الحية، ويُمْكِن توظيفُها كلغةٍ ارتكازيةٍ في نُظُم الترجمة الآلية بين أزواج اللغات المختلفة، وكذلك كلُغةٍ للتمثيل المعرفي العام تستفيدُ منها على سبيل المثال أنظمة استرجاع المعلومات.

وقد نشأت هذه اللغة عامَ ١٩٩٦م في معهد الدراسات المتقدمة بجامعة الأمم المتحدة في «طوكيو باليابان» وتم النشرُ عنها لأول مرة عام ١٩٩٩م، ويَسْتَمِرُّ تطويرُها منذ ذلك الحين عبر شبكة عالمية واسعةِ النطاق من الباحثين والمطوِّرين، كما أُنْشِئَتْ لذلك مؤسَّسةٌ غيرُ هادفةٍ للربح تحت اسم (UNDL) وهي مُشْهَرَةٌ في «جِنِيف بسويسرا» في عام ٢٠٠١م للقيام على هذا المشروع العملاق، وتعلن هذه المؤسسة أن هذا العمل مِلْكٌ للأمم المتحدة مما يعني أنه مِلْكٌ لعموم البشرية.

وتُعَدُّ «لغةُ الشبكات الدِّلالية الحاسوبية العالمية» مثالاً عميقًا ومفصَّلاً في اتجاه التمثيل المعرفي اللغوي البشري المشترَك المستقل عن خصوصياتِ كُلِّ لغةٍ حية، حيث جرى تصميمُ طبقات ومكوِّنات هذه اللغة بالتناظر مع العناصر الرئيسية في اللغة الحية من مفرَداتٍ ونحوٍ وأنطولوجيا ولكن بصورة حيادية[1]. وقد أنجز هذا المشروعُ العملاقُ بِنْيَتَه الأساسية عبر السنين الماضية بواسطة جهود مئتي باحث مُكَرَّسٍ لذلك فضلاً عن الآلاف غيرهم ممن يساهمون ويَنْشُرون حول هذا المشروع عبر أرجاء الأرض، وما زال المشروع قائمًا على تحسين هذه البنية التحتية كما لا يزال يواصل إنجاز الجسور بين اللغات الحية الكبرى (ومنها العربية) إلى هذه اللغة الحيادية.

ولا تزال هناك تحديات أساسية تواجه هذا المشروع؛ منها على سبيل المثال الالتباس الكبير على عدة مستويات لغوية أثناء النقل من أية لُغَةٍ حيةٍ إلى لغة الشبكات الدلالية الحاسوبية العالمية (الحيادية)، ومنها أيضًا قصورُ قدرات هذه اللغة الحيادية عن التعبير بشكل كامل عن الدِّلالات والمقصودات العميقة لعبارات النصوص في لغاتها الأصلية!

١- مع مراعاة استبعاد عناصر الالتباس الكامنة في اللغات الحية من هذه التصميمات.

كما يقع في هذا الإطار مشروعٌ آخَر تحت اسم «أُونْطُو- نُوتْس» (OntoNotes) [٣٦] لكنه على جانب بناء الموارِد اللغوية اللازمة لتدريب الأنظمة «المرْتقَبة» للمعالجة الحاسوبية الدِّلالية العميقة لِلُّغة الحية عَبْرَ العنونة الدِّلالية-النحوية للمدونات النصية على مستوى الجُمَل. وقد انطلق «أُونْطُو-نُوتْس» خلال العام ٢٠١٠م/ ٢٠١١م وسيظل تحت التنفيذ لمدة خمس سنوات وتقوم عليه شركةُ «بي-بي-إن» (BBN) مع جامعة كولورادو، وجامعة بِينْسِلْفَانْيَا، ومعهد علوم المعلومات بجامعة جنوب كاليفورنيا، وتقع كل هذه المؤسسات في الولايات المتحدة الأمريكية.

ويستهدف المشروعُ إتمامَ العنونة اليدوية لمدوَّناتٍ نَصِّيةٍ بالإنجليزية (مليون كلمة) وبالصينية (مليون كلمة) وبالعربية (نصف مليون كلمة)، وهي تغطي نطاقاتٍ متنوِّعة {أخبار، محادَثات تليفونية، تدوينات إلكترونية «بلُوجْز» (Blogs)، نشرات أخبار، محاضَرات ...} وذلك عبر عنونةٍ ذاتِ دقةٍ عاليةٍ جدًّا واتفاقٍ يجاوز تسعين في المئة بين مختلف المعَنْوِنين.

والعنونةُ المستهدَفةُ هي عنونة تركيبية (نحوية-دِلالية) على مستوى الجُمَل الكاملة؛ ترتبط فيها المعلوماتُ النحويةُ (إعرابية + تراكيب «مُسْنَد - مُسْنَد إليه») بالمعلومات الدِّلالية السطحية (التي تربطُ كُلَّ كلمةٍ بالمعنى الذي تَرِدُ به في أنطولوجيا، وكذلك بعائدٍ يُنْسَبُ إليه المعنى). وسوف يتيحُ المشروعُ بعد إتمامه هذه المدونات المعنوَنة للباحثين بصورةٍ مفتوحة المصدر من أجل تشجيعهم على تطبيق خُوارِزْمات التعلُّم الحاسوبي عليها سعيًّا إلى تدريبها على تحليل وإنتاج هذه الارتباطات والتراكيب حاسوبيًّا في وثائق جديدة خلاف تلك التي أنتجها هذا المشروع.

- لا تزال المعالجة الحاسوبية لِلُّغة دِلاليًّا فيما يتجاوز المستوى المعْجَمي تحديًا ضخمًا، وتُعَدُّ الآن الحقلَ الأكثرَ خصوبةً في إطار مُعالجَة اللغة حاسوبيًّا الذي ينشط فيه الباحثون من الحاسوبيين واللُّغَويين معًا، وفي هذا الصدد تحتاج اللُّغةُ العربيةُ على وجه الخصوص إلى تكريس العديد من مثل هؤلاء الباحثِين لجهودٍ غزيرةٍ على مدى طويل من أجل:
- الاتفاقِ أولاً على الصياغات الرياضية المحكمة للمسائل الدلالية المركزية على المستويات فوق المعجمية، ثم إيجادِ حلول رياضية مقبولة لها، ثم ترويضِ التكلفة الحاسوبية لهذه الحلول حتى تصيرَ في متناول قدرات الحاسبات الرقمية المعاصرة ويمكنَ بذلك أخيرًا إدماجُها في تطبيقاتٍ عمليةٍ ضمن إطار تقنيات معالجة اللغات الحية.
- العمل على العنونة الدلالية المعجمية – وكذلك العنونة الدلالية فوق المعجمية – لمدوَّناتٍ نصيةٍ متوازنةِ النطاقات وذات أحجام كبيرة وبدقة عالية، وهو أمر تمس الحاجة إليه من أجل تدريب مختلف آليات فك الالتباس الدلالي.

ببليوجرافيا مرجعيَّة

١. أنيس (إبراهيم): دلالة الألفاظ، مكتبة الأنجلو المصرية، ١٩٥٢م.

٢. جبل (عبد الكريم حسن): في علم الدلالة، دار المعرفة الجامعية، الإسكندرية، ١٩٩٧.

٣. صيني (محمود إسماعيل) وآخرون: المكنز العربي المعاصر، مكتبة لبنان، بيروت، ط١، ١٩٩٣م.

٤. عُمَر (أحمد مُختار) وآخرون: المَكْنَزُ الكَبيرُ، دارُ نَشْرِ «سُطُور» المملكة العربية السعودية، ط١، ٢٠٠٢م.

٥. عُمَر (أحمد مُختار): عِلْمُ الدِّلَالةِ، عالَمُ الكُتُبِ، ط٥، ١٩٩٨م.

٦. غالب (حنَّا): كنز اللغة العربية، لبنان ناشرون، ٢٠٠٣م.

٧. غالي (وجدي رزق): معجم المترادفات العربية الأصغر، مكتبة لبنان، بيروت،

ط١، ١٩٩٦م.

٨. فَيَّاض (سُلَيْمان): الحُـقولُ الدِّلَالِيَّةُ الصَّرْفِيَّةُ لِلأَفْعالِ العَرَبيَّةِ، دارُ المرِّيخ بالرِّياض، ١٩٩٠م.

٩. مدينة الملك عبد العزيز للعلوم والتقنية – بالاشتراك مع – المنظمة العربية للتربية والثقافة والعلوم، المعجم العربي التفاعلي؛ مواصفات وخصائص المشروع، يونيو ٢٠٠٨م. http://www.almuajam.org/AraDicPlan_3.pdf

١٠. المنظمة العربية للتربية والثقافة والعلوم، المعجم العربي الأساسي، لاروس، ١٩٨٨م.

١١. نخلة (رفائِيل اليَسوعيّ): المُنْجِد في المترادفات والمتجانسات، دار المشرق.

١٢. اليازجي (إبراهيم): نجعة الرائد في المترادف والمتوارد، مكتبة لبنان، بيروت.

13. Agirre, E., Edmonds, P., Word Sense Disambiguation; Algorithms and Applications, Springer-Verlag, ISBN 978-1-4020-4808-4, 1st ed. 2007.

14. Attia, M., Rashwan, M., Al-Badrashiny, M., Fassieh©; a Semi-Automatic Visual Interactive Tool for the Morphological, PoS-Tags, Phonetic, and Semantic Annotation of the Arabic Text, IEEE Transactions on Audio, Speech, and Language Processing (TASLP) http://www.SignalProcessingSociety.org/Publications/Periodicals/TASLP: Special Issue on Processing Morphologically Rich Languages, Vol. 17 - Issue 5; pp. 916 to pp. 925 http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?isnumber=5067414& ar-number=5075778&count=21&index=6, July 2009.

15. Attia, M., Rashwan, M., Ragheb, A., Al-Badrashiny, M., Al-Basoumy, H., Abdou, S., A Compact Arabic Lexical Semantics Language Resource Based on the Theory of Semantic Fields, Lecture Notes on Computer Science (LNCS): Advances in Natural Language Pro-cessing, Springer-Verlag Berlin Heidelberg: www.SpringerOnline.com, LNCS/LNAI; Vol. No. 5221/2008; pp. 65 – 76, http://www.springerlink.com/content/100p13145723v162/, Aug. 2008.

16. Black, W., Elkateb, S., Rodriguez, H., Alkhalifa, M., Vossen, P., Fell-baum, C., Introducing the Arabic Word Net Project, 2006; http://www.globalwordnet.org/AWN/meetings/GWApaper.pdf.

17. Diab, M., The Feasibility of Bootstrapping an Arabic Word Net Lev-eraging Parallel Corpora and an English Word Net, Proceedings of the Arabic Language Technologies and Resources Int'l Conference; NEMLAR, Cairo 2004.

18. Dichy, J., Hassoun, M., The DINAR.1 (DIctionnaire INformatisé de l'ARabe, version 1) Arabic Lexical Resource, an Outline of Contents and Methodology, The ELRA news letter, April-June 2005, Vol.10 n.2, France.

19. Ghonaimy, M.A., A Tutorial Review on Word Nets, Proceedings of the 4th Conference on Language Engineering; CLE'2003, the Egyptian Society of Language Engineering (ESoLE).

20. Gruber, T., What is an Ontology? http://www-ksl.stanford.edu/kst/what-is-an-ontology.html, Stanford University, 2001. (Retrieved Nov. 9th, 2009(.

21. Gruber, T., Toward Principles for the Design of Ontologies Used for Knowledge Sharing, International Journal of Human-Computer Studies 43 (5-6): 907–928, 1995.

22. Hearst, M., Untangling Text Data Mining, Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics (ACL), 1999; http://www.sims.Berkeley.edu/~hearst/papers/acl99/acl99-tdm.html.

23. Kahusk, N., Vider, K., Lexical Semantic Databases, Summer School Language Technology in Human-Computer Interaction, Institute of Computer Science - Department of General Linguistics - University of Tartu - Estonia, Aug. 2002. http://www.cs.ut.ee/~koit/SS02/lex-sem_over1.pdf.

24. Lehrer, A., Semantic Fields and Lexical Structures, Amsterdam - London, 1974.

25. Riloff, E., Jones, R., Learning Dictionaries for Information Extraction Using Multi-level Boot-strapping, Proceedings of AAAI-99.
26. Schütze, H., Manning, C.D., Foundations of Statistical Natural Language Processing, the MIT Press, 2000.
27. Vossen, P., Euro Word Net; General Document, Version 3 - Final, University of Amsterdam, http://www.hum.uva.nl/~ewn, 2002.

مواقع ذات صلة على الشبكة العنكبوتية:

28. http://en.wikipedia.org/wiki/Ontology_(information_science)
29. http://www.lT-world.org/
30. http://www.globalwordnet.org/
31. http://www.globalwordnet.org/AWN
32. WordNet® - Princeton University: http://wordnet.princeton.edu/
33. http://arabicontology.sourceforge.net/
34. http://www.undl.org/
35. http://www.rdi-eg.com/ar/technologies/arabic_nlp.htm,
36. http://www.rdi-eg.com/Technologies/arabic_nlp.htm,
37. http://www.rdi-eg.com/ar/projects/TextMining.htm
38. http://www.MonNet-project.eu
39. http://www.BBN.com/NLP/OntoNotes
40. http://kacst.summon.serialssolutions.com/

الفصل السَّادس
التَّحليل المُعجميّ الآليّ

د. الـمُعتزّ بالله السَّعيد

١- في التَّحليل المُعجميّ.
٢- من مُكوِّنات المُعجم اللُّغَويّ.
٣- المُعالَجة المُعجميَّة الآليَّة في مراحل الصِّناعة.
٤- من أدوات المُعالَجة الآليَّة لمُكوِّنات المُعجم اللُّغَويّ.
٥- أفكارٌ بحثيَّة لأطرُوحاتٍ علميَّةٍ ودراسات مُستقبليَّة.

١ – في التَّحليل المُعجميّ

يُعنى التَّحليلُ الـمُعجميُّ بمُعالَجَة مُكوِّنات الـمُعجم اللُّغويّ الَّتي تنبثقُ عن رُكنَين رَئيسَين، هُما: المبنى، والمعنى. ومن هذا الـمُنطَلَق، فإنَّ التَّحليلَ الـمُعجميَّ يتناولُ أساليبَ البحث في الثَّروة اللَّفظيَّة للُّغة الإنسانيَّة، وطُرُق جمع مُفرَدات اللُّغة وتصنيفها وترتيبها وتعيين صِيَغِها ومُشتقَّاتِها ومعانيها وأساليبها، بالإضافة إلى شواهدِها الدَّالَّة على استعمالاتِها اللُّغويَّة.

ويُمثِّلُ المبنى في المُعجم اللُّغويّ الوَحدة المفتاحيَّة الَّتي يُستَنَدُ إليها في الاستدلال على المعارف المُعجميَّة (من المعاني، والشَّواهد، والاستعمالات، ...)؛ ويُعَبَّرُ عنهُ في الصِّناعة المُعجميَّة الحديثة بما يُعرَفُ بـ (لكسيم Lexeme) أو (الوَحدة المُعجميَّة) أو (وَحدة التَّحليل المُعجميّ). وفي اللُّغات الاشتقاقيَّة مثل اللُّغة العربيَّة تتفرَّعُ الوَحدةُ المُعجميَّةُ عن أصلٍ [أو جذر] للكلمة، يُعرَفُ بـ (المدخل المُعجميّ)، ويُمكنُ تعريفُهُ بأنَّهُ الحَقل الَّذي تنتَمي إليه مجموعةٌ من الكلمات الَّتي تشتركُ في مادَّةٍ لُغويَّةٍ واحِدة؛ وسيردُ تفصيلُ ذلكَ لاحقًا.

أمَّا المعنى فهو المُكوِّنُ المُستَهدَفُ الَّذي يُصنَعُ المُعجمُ أساسًا لأجل الوصول إليه؛ ويُمثِّلُ المفهومَ اللُّغويَّ المُجتَمَعيَّ لمُفرَدات اللُّغة المُمثَّلة في الوحدات المُعجميَّة. وهو بذلكَ وسيلةُ إزالة التباسِ المُفرَدة؛ ويُعتَمَدُ في استنباط المعاني المُعجميَّة على الدِّلالات اللُّغويَّة لمُفرَدات اللُّغة في السِّياقات الَّتي تردُ فيها؛ وتبعًا لذلكَ، يُحتَمَلُ أن تتعدَّدَ المعاني المُعجميَّة للمُفرَدة الواحدة بحسب سياقاتها. وبطبيعة الحال، فإنَّ هذه السِّياقات تُستَمَدُّ من اللُّغة الطبيعيَّة المُستَخدَمة فعليًّا.

يُعنى التَّحليلُ المُعجميُّ إذن بمُكوِّنات المُعجم اللُّغويّ الَّتي تبدأ بالمداخل؛ وتتفرَّعُ عن كلِّ مدخلٍ وَحدةٌ أو مجموعةٌ من الوحدات المُعجميَّة؛ ويتفرَّعُ عن كلِّ وَحدةٍ معنًى أو مجموعة من المعاني المُعجميَّة، ويتَّصلُ كُلُّ مبنًى بمعانٍ وظيفيَّة، ويتَّصلُ كلُّ معنًى بشواهدَ ومعارفَ مُعجميَّة أخرى، وهكذا. وسنعرضُ لبيان هذه المُكوِّنات ودَورها في المُعجم اللُّغويّ فيما يأتي.

٢- من مُكوِّنات المُعجم اللُّغَويّ

تخضعُ بنيةُ المُعجم اللُّغويّ لعدَّة عوامل، منها: طبيعةُ اللُّغة، والغَرَضُ من المُعجم، والفئاتُ المُستهدَفة. لهذا، فإنَّ المعاجمَ اللُّغويَّةَ تأخذُ أنماطًا مُختلفة. وبطبيعة الحال، فإنَّ هذا الاختلافَ يُؤدِّي إلى تبايُن مُكوِّنات المُعجم [أو عناصره]. وعلى سبيل المثال، فإنَّ المعاجمَ المُعاصرة لا تُعنى بمعلومات التَّأثيل الَّتي تُعَدُّ مُكوِّنًا رئيسًا في المعاجم التَّاريخيَّة؛ والمعاجمُ المُوجَّهة لأبناء اللُّغة المُعيَّنة لا تُعنى بمعلومات التَّكرار الَّتي تُعنى بها المعاجمُ التَّعليميَّة المُوجَّهة للنَّاطقينَ بغير هذه اللُّغة، وقِس على ذلك أنواعَ المعاجم باختلاف لُغاتِها ومُستوياتها وأهدافها.

يُمكنُ القولُ إنَّ هناكَ مُكوِّناتٍ مُعجميَّةً رئيسة، لا يستقيمُ المُعجمُ بدونها؛ ولا يكونُ مُعجمًا بالمفهوم الدَّقيق إلَّا بوُجودها، مثل: (الوحدات المُعجميَّة) الَّتي تُعَبِّرُ عن «اللُّغة الموصوفة»؛ وهي المُمَثَّلة في المُفرَدات [الغامضة] الَّتي ينشُدُ صُنَّاعُ المُعجم شرحَها وإزالَةَ الالتباس عنها، وكذلكَ المعاني المُعجميَّة الَّتي تُعَبِّرُ عن «اللُّغة الواصفة»؛ وهي مادَّةُ الإفهام الَّتي تُؤدِّي حاجةَ مُستخدمي المُعجم وتُحقِّقُ الهدفَ الأساسيَّ من الصِّناعة المُعجميَّة.

من هذا المُنطَلَق، يُمكنُ تصنيفُ مُكوِّنات المُعجَم وفقَ اعتبارَين:

- **اعتبار العُمُوم والاختصاص:**

حيثُ تصلحُ بعضُ المُكوِّنات لأن تكونَ عُنصرًا في مُختَلَف أنواع المعاجم، وتصلحُ مُكوِّناتٌ أخرى لأن تكونَ عُنصُرًا في معاجمَ مُختَصَّةٍ مصنوعةٍ لأهدافٍ مُعَيَّنة.

- **اعتبار الثَّبات والتَّغيُّر:**

ونعني أن تأخُذَ بعضُ المُكوِّنات شكلًا ثابتًا تفرضُهُ قواعدُ اللُّغة وقوانينُها، وأن تأخُذَ مُكوِّناتٌ أخرى شكلًا مُتغيِّرًا تتحكَّمُ فيه المادَّةُ المُعجميَّةُ ذاتُها.

ويُوضِّحُ (الجدول ٦-١) بيانًا لأبرز مُكوِّنات المُعجم اللُّغويّ، وحالتها بينَ العُمُوم والاختصاص من ناحية، والثَّبات والتَّغيُّر من ناحيةٍ أخرى.

م	مُكوِّنات المُعجم اللُّغويّ	العُمُوم والاختصاص		الثّبات والتَّغيُّر	
		عامَّة	مُختَصَّة	ثابتة	مُتغيِّرة
١	المداخل والوحدات المُعجميَّة	+		+	
٢	معلومات الهجاء والنُّطق	+		+	
٣	المعاني الوظيفيَّة	+		+	
٤	المعاني المُعجميَّة	+		+	
٥	المعلومات التَّأثيليَّة		+	+	
٦	الحُقُول والعلاقات الدِّلاليَّة		+	+	
٧	الشَّواهد المُعجميَّة	+			+
٨	المعلومات التَّاريخيَّة		+		+
٩	مُستويات الاستعمال		+		+
١٠	معلومات التَّكرار [التَّرَدُّد]		+		+

الجدول ٦-١: نماذج من الوحدات المعجميَّة في الـمُعجم العربيّ

ونعرضُ فيما يأتي لماهيَّة هذه المُكوِّنات، كُلٍّ على حِدة، مع التَّمثيل عليها في المُعجم العربيّ.

٢, ١- المداخل والوحدات المُعجميَّة (Entries & Lexemes)

المدخلُ المُعجَمِيُّ Lexical Entry هو ذلك الحَقلُ الَّذي تنتَمي إليه مجموعةٌ من الكلِمات الَّتي تشترِكُ في مادَّةٍ لُغَوِيَّةٍ واحِدة [في معاجم الألفاظ]، وقد يكونُ جذرًا لُغَوِيًّا لكلمةٍ عَرَبِيَّةٍ أو مُعَرَّبة، أو يكونُ مادَّةً مُعجَمِيَّةً [مجموعة من المحارف الأصيلة] لكلمَةٍ دخيلة؛ أمَّا الوحدات المُعجَمِيَّة Lexemes فهي مجمُوعَةُ الكَلِمات الرَّأسِيَّة الَّتي تُشَكِّلُ القَوائِمَ المُنسَدِلَةَ عن المَدخل.

ويُوضِّحُ (الجدول ٦-٢) نماذجَ المداخل والوحدات في المُعجم العربيّ.

م	الوحدات المُعجميَّة	المداخل المُعجميَّة	أقسام الكلام	أنواع الوَحدات
١	آثَرَ		فِعل	عربيَّة
٢	استأثَرَ	ء ث ر	فِعل	عربيَّة
٣	الأثَر		اسم	عربيَّة
٤	الباشا	ب ا ش ا	اسم	دخيلة
٥	الأُسْقُفّ	س ق ف	اسم	مُعَرَّبة
٦	السَّقْف		اسم	عربيَّة
٧	كَتَبَ		فِعل	عربيَّة
٨	الكاتِب	ك ت ب	اسم	عربيَّة
٩	المكتَبَة		اسم	عربيَّة
١٠	يا	ي ا	حرف / أداة	عربيَّة

الجدول ٦-٢: نماذج من الوحدات المعجميَّة في الـمُعجم العربيّ

٢ ,٢- معلومات الهجاء والنُّطق (Alphabet & Pronunciation)

يُقصَدُ بها المعلوماتُ الَّتي تُوجِّهُ مُستخدمي المُعجم إلى معرفة طريقة كتابة الوحدات المُعجميَّة [أو المُفرَدات المشروحة] وكيفيَّة نُطقِها بصورةٍ سليمة، وفقًا لقواعد اللُّغة. والهدفُ من هذه المعلومات تعليميٌّ في الأساس؛ حيثُ تُساعدُ مُتعلِّمي اللُّغة على الإفادة من المُعجم في تعلُّم المهارات اللُّغويَّة: القراءة، والكتابة، والتَّحدُّث والاستماع.

فيما يتعلَّقُ بمعلومات الهجاء، نُلاحظُ عنايةَ المعاجم العربيَّة القديمة بها؛ حيثُ كانت تنُصُّ على أنَّ كلمةً ما تُكتَبُ بالعَين المُهملةِ تمييزًا لها عن الغَين، وأنَّ كلمةً ثانيةً تُكتَبُ بالباء المُوَحَّدة تمييزًا لها عن الياء، وأنَّ كلمةً ثالثةً تُكتَبُ بالتَّاء المُثنَّاة تمييزًا لها عن الثَّاء الَّتي تُوصَفُ بأنَّها مُثلَّثة، وهكذا. ومع تطوُّر شكل الكتابة العربيَّة وظهور النِّقاط وعلامات الضَّبط، تقلَّصَت هذه العناية، حتَّى تلاشَت في أكثر المعاجم المُعاصرة.

أمَّا معلوماتُ النُّطق، فيغلُبُ على المعاجم العربيَّة أن تكتفي بوضع علامات الضَّبط لِفبانة عن النُّطق السَّليم؛ إلَّا أنَّ ثمَّةَ بعضَ المعاجم الَّتي تُعنى بإبراز طريقة نُطق الوحدات المُعجميَّة باستخدام رُموز الألِفبائِيَّة الصَّوتِيَّة الدَّولِيَّة International Phonetic Alphabet (IPA)؛ وهي مجموعةٌ من رُموز المحارف، أقرَّتها الجَمعِيَّة الصَّوتِيَّة الدَّولِيَّة

بباريس Association Phonétique Internationale في عام ١٨٨٨م، لتيسير قراءة مُختَلَف اللُّغات الإنسانيَّة. ويُوضِّحُ (الجدول ٦-٣) تصوُّرًا لرُموز الكتابة الصَّوتيَّة للفونيمات [الأصوات] العربيَّة.

الصَّوت	الرَّمز		الصَّوت	الرَّمز		الصَّوت	الرَّمز
الصَّوامت وأشباه الصَّوامت Consonants & Semi Consonants							
ء	ʔ		ر	R		ف	f
ب	B		ز	Z		ق	q
ت	T		س	S		ك	k
ث	ṯ		ش	Š		ل	l
ج	dʒ		ص	ṣ		م	m
ح	Ħ		ض	ḍ		ن	n
خ	ḫ		ط	ṭ		هـ	h
د	D		ظ	ẓ		و	w
ذ	ḏ		ع	ʕ		ي	y
			غ	Ġ			
الصَّوائت Vowels							
(ا) قصيرة	A		(و) قصيرة	ʊ		(ي) قصيرة	i
(ا) طويلة	Æ		(و) طويلة	u:		(ي) طويلة	i:

الجدول ٦-٣: تصوُّر لرُموز الكتابة الصَّوتيَّة للوَحَدات المُعجميَّة العربيَّة

٢,٣- المعاني الوظيفيَّة (Grammatical Functions)

يُقصَدُ بها الوظائف النَّحويَّة الَّتي تُؤدِّيها مُفرداتُ المُعجم في سياقاتٍ لُغويَّةٍ تتجاوزُ إطارَ الكلمة الواحدة إلى تراكيب واستعمالاتٍ لُغويَّة. وتقتصرُ المعاجمُ اللُّغويَّةُ عمومًا على المعاني الوظيفيَّة السَّماعيَّة دونَ القياسيَّة. وتأتي المعاني الوظيفيَّةُ في المُعجم العربيّ على إحدى صُورَتَين:

- (**المعاني الوظيفيَّة البِنَويَّة**). تُعنى بالبنية الصَّرفيَّة للمُفرَدات؛ ومن أمثلتها: ما يتَّصلُ بأبواب الأفعال الثُّلاثيَّة المُجرَّدةِ، ومصادرَها، وصِيَغ التَّذكير والتَّأنيث، وجُمُوع التَّكسير.

- **(المعاني الوظيفيَّة التَّركيبيَّة)**. تُعنى بالبنية التَّركيبيَّة؛ ومن أمثلتها: التَّمييز بينَ أقسام الكلام العربيّ (الاسم والفعل والأداة)، والتَّمييز بينَ اللَّازم والمُتعدِّي من الأفعال.

والدَّاعي إلى وُجود هذه المعاني في المُعجم أنَّها لا تخضعُ لقاعدةٍ نحويَّةٍ مُعَيَّنةٍ يُمكنُ الاهتداءُ بها أو القياسُ عليها، وليسَت ممَّا تُعنى به مصادرُ التَّقعيد النَّحويّ للُّغة. ومن ثَمَّ، تبدو الحاجةُ إليها لمُساعدة مُستخدمي المُعجم على توظيف المُفرَدات واشتقاقاتِها في استعمالاتٍ لُغويَّةٍ سليمة. وعلى سبيل المثال، تُساعدُ المعاني الوظيفيَّة البنويَّة على توجيه مُستخدمي المُعجم إلى صيغة الفعل المُضارع (يَكتُب) من الماضي الثُّلاثيّ المُجرَّد (كَتَب)؛ و تُساعدُ المعاني الوظيفيَّة التَّركيبيَّةُ على توجيه المُستخدمينَ إلى أنَّ الفعلَ (أعطى) لا يستقيمُ في الجُملة العربيَّة إلَّا بوُجود المفعولَين، بمعنى أنَّهُ فعلٌ مُتعدٍّ بمفعُولَين؛ وهكذا.

٢ , ٤ – المعاني المُعجميَّة (Lexical Meanings)

المعنى المُعجميُّ هو العاملُ المُشترَكُ بينَ المُعجماتِ اللُّغَويَّة على اختلافِ أنواعها؛ إذ يُعَوَّلُ عليه أساسًا في تحقيق غاية هذه المُعجمات من الإفهام وإزالة الغُمُوض؛ ويُقصَدُ به ذلكَ المعنى الَّذي تُشيرُ إليه الوحداتُ المُعجميَّة في سياقاتِها اللُّغويَّة، شريطةَ أن تكونَ هذه السِّياقاتُ مُعَبِّرةً عن العُرف العامّ لدى الجماعة اللُّغويَّة، أو بعبارةٍ أخرى، شريطةَ أن تكونَ الاستعمالاتُ اللُّغويَّة لهذه السِّياقات شائعةً ومقبولةً ومُتعارَفًا عليها في المُجتمع اللُّغويّ المُعَيَّن.

ويأتي المعنى المُعجميُّ على إحدى صُورَتَين:

- **المَعنى الحَقِيقِيّ Literal meaning**: وهو المعنى الصَّريح المُباشر الَّذي تُشيرُ إليه مدلولاتُ الوحدات المُعجميَّة.
- **المَعنى المَجازِيّ Figurative meaning**: وهُو المعنى غَير المُباشر الَّذي يحملُ دلالةً مُغايرةً للمعنى الحقيقيّ؛ ويكونُ هذا المعنى ذا طابعٍ بلاغيٍّ، وَتَدُلُّ عليه قَرِينةٌ لَفظِيَّةٌ أو مَعنَوِيَّة.

وَتَتَعَدَّدُ المَعانِي المُعجَمِيَّةُ للوحدات [المُفرَدات] بِتَعَدُّد السِّياقات وأساليب التَّعبير الَّتي تردُ فيها المُفرَدة. ويُمكنُ التَّمثيلُ على ذلكَ بالوَحدة المُعجميَّة (العَين)؛ حيثُ تردُ أحيانًا فتُفيدُ «عُضو الإبصار عند الإنسان والحيوان»، وقد يُرادُ بها «يَنبُوع الماء الَّذي

يَجرِي في الأرض»، وتَرمِي – مجازًا – إلى «الجاسوس»، وإلى «طَلِيعَة الجَيش»، وإلى «كَبِير القَوْم»، وإلى «ذات الشَّيء»، وقد تَرمِي إلى مَعنًى مَجازِيٍّ آخَر يُفِيدُ «الإكرامَ والحِفْظ» كأن تَقُول – على سبيل الوُدّ والأُلفة (أنتَ على عَينِي)، وغيرِها مِن المعاني الَّتي تتَضِّحُ من خِلال السِّياق. ومثلُ ذلكَ نجدُهُ في الوَحدة المُعجميَّة (الحاجب)؛ حيثُ تُشيرُ إلى (الشَّعْر النَّابِت فَوقَ العَيْن) وتُشيرُ – أيضًا – إلى (خازن الباب وحارِسه)؛ وكذلكَ الوحدة (القانون) الَّتي تُشيرُ إلى (الآلة الموسيقيَّة الوتريَّة) وإلى (قَواعِدَ وأحْكَام يَتَّبِعُها النَّاسُ في علاقاتِهِم المُخْتَلِفَة وتُنَفَّذُ بواسِطة المحاكِم).

٢, ٥- المعلومات التَّأثيليَّة (Etymological Information)

يُعنى التَّأثيل المُعجميّ Lexical Etymology بتعيين المعلومات الَّتي تبحثُ في أصول الوحدات المُعجميَّة في مرحلة ما قبلَ دُخول نُصُوص اللُّغة، مع تتبُّع التَّطوُّر اللُّغويّ الحادث في هذه الأصول، حتَّى مرحلة الاستقرار في نُصُوص اللُّغة المُعَيَّنة.

وقد اهتمَّ صُنَّاعُ المعاجم بالتَّأثيل المُعجميّ في العديد من اللُّغات الطَّبيعيَّة، مثل: الألمانيَّة والإنجليزيَّة والفرنسيَّة، فصَنَعوا معاجمَ مخصوصةً لذلك، عُرِفَت بالمعاجم التَّأثيليَّة. أمَّا في اللُّغة العربيَّة، فقد ظهرَ التَّأثيلُ المُعجميُّ في حقبةٍ زمنيَّةٍ مُتأخِّرةٍ نسبيًّا، في بعض تجارب الصِّناعة المُعجميَّة غَير المُكتملة، مثل «المُعجم الكبير» الَّذي يعملُ على إنجازِهِ مجمعُ اللُّغة العربيَّة بالقاهرة، و «مُعجم الدَّوحة التَّاريخيّ» الَّذي يُشرفُ عليه المركز العربيّ للأبحاث في الدَّوحة. وتعتمدُ مثلُ هذه المعاجم في تأثيل المُفرَدات على تعيين النَّظائر المُشابهة لمبنى الكلمة العربيَّة ومعناها في اللُّغات السَّاميَّة، أو لُغات بعض الفصائل اللُّغويَّة الَّتي كانت على صلةٍ باللُّغات السَّاميَّة.

٢, ٦- الحُقُول والعلاقات الدِّلاليَّة (Semantic Fields & Semantic Relations)

يُقصَدُ بالحُقول الدِّلاليَّة Semantic Fields المجالات الموضوعيَّة [الدِّلاليَّة] الَّتي تتبعُها مجموعةٌ من المُفرَدات المُنتمية إلى حقلٍ مُعَيَّن، مثل حقل (الملبَس) الَّذي يضُمُّ (البنطالَ، والجلبابَ، والعمامة). وإذا كانَ الجذرُ اللُّغويُّ [وما شاكلَهُ من الأصول] يُمثِّلُ صُورةَ المدخل المُعجميّ في معاجم الألفاظ، فإنَّ الحُقُولَ الدِّلاليَّة تُمثِّلُ أوعيةَ المُفرَدات في معاجم المعاني [الموضوعات] الَّتي تجمعُ مُفرَدات اللُّغة وتُرتِّبُها انطلاقًا من المعنى.

أمَّا العلاقاتُ الدِّلاليَّة Semantic Relations فهي وسيلةُ الرَّبط بينَ الوحدات المُعجميَّة. وتأتي هذه العلاقاتُ في المُعجم اللُّغويّ على إحدى صُورَتَين:

- **العلاقات الأفقيَّة:** تكونُ بينَ مجموعةٍ من المُفرَدات الَّتي تقعُ في مُستوى دلاليٍّ مُتوازٍ؛ مثل علاقات (التَّرادُف، والتَّنافُر، والتَّضادّ، والاشتراك اللَّفظيّ)
- **العلاقات الرَّأسيَّة:** تكونُ بينَ مجموعةٍ من المُفرَدات الَّتي تقعُ في مُستوى دلاليٍّ مُتتالٍ [عموديّ]؛ مثل علاقات (الاشتمال، والنَّوعيَّة، والكُلِّيَّة، والجزئيَّة).

وتجدُرُ الإشارةُ إلى عناية المُعجم العربيّ بالموضوعات في وقتٍ مُبكِّر؛ حيثُ ظهرَت العديدُ من المُصَنَّفات الَّتي اعتنَت بالحُقُول والعلاقات الدِّلاليَّة، مثل: (الغريب المُصَنَّف) للقاسم بن سلَّام (٢٢٤هـ)، و (جواهر الألفاظ) لقُدامة بن جعفر (٣٧٧)، و (فقه اللُّغة وسرّ العربيَّة) لعبد الملك بن مُحمَّد الثَّعالبي (٤٢٩)، و (المُخَصَّص) لابن سِيدَة الأندلُسيّ (٤٥٨هـ)، وغيرها.

٢,٧- الشَّواهد المُعجميَّة (Lexical Evidence)

تُعَدُّ الشَّواهدُ المُعجميَّةُ دليلًا على وُجود استعمالٍ لُغويٍّ حقيقيٍّ للوحدات المُعجميَّة ومعانيها؛ وتُمثِّل بذلك إحدى وسائل شرح المعنى، لا سيَّما في الوحدات والمعاني المُعجميَّة الَّتي تحملُ دلالاتٍ غير دائرة في المُجتَمَع اللُّغويّ. وتستمدُّ الشَّواهدُ المُعجميَّةُ مادَّتها في الغالب من (المُدوَّنات اللُّغويَّة Linguistic Corpora)؛ وهي مجموعاتٌ كبيرةٌ نسبيًّا من نُصُوص اللُّغة المُمثِّلة للواقع اللُّغويّ، والمُستمدَّة منه أساسًا. وفي بعض الأحيان، يلجأ صُنَّاعُ المعاجم إلى وضع أمثلةٍ مصنوعةٍ لتقومَ مقامَ الشَّواهد المُعجميَّة.

والواقعُ أنَّ اختيارَ الشَّواهد المُعجميَّة لا يتمُّ بصورةٍ عشوائيَّة؛ إذ ينبغي أن تتحقَّقَ معهُ غايةُ الإفهام وإزالة التباس المباني والمعاني. لهذا، ينبغي إخضاعُ الشَّواهد المُعجميَّةُ لبعض الضَّوابط المنهجيَّة الَّتي تضمنُ تحقيقَ الغاية. ومن هذه الضَّوابط، على سبيل المثال: السَّلامة اللُّغويَّة، والإيجاز، والوُضُوح، وشُيوع البنية التَّركيبيَّة للشَّاهد في المُجتَمَع اللُّغويّ.

٢,٨- المعلومات التَّاريخيَّة (Historical Information)

تُعنى المعلوماتُ التَّاريخيَّةُ في المعاجم اللُّغَويَّة بتعقُّب التَّطوُّرات والتَّغيُّرات الحادثة في أشكال المباني ومعانيها منذُ أقدم ظُهورٍ مُؤَرَّخٍ لها في صُورتِها المكتوبة [المعلومة لدى أهل

اللُّغة]؛ كما تبحثُ في المُستعمَل والمُهمَل والمُمات من مُفرَدات اللُّغة عبرَ تاريخِها.

والواقعُ أنَّ التَّعاطي مع هذه المعلومات ليسَ أمرًا يسيرًا في المعاجم العربيَّة؛ حيثُ يفرضُ تاريخُ اللُّغة العربيَّة وُجُودَ قدرٍ كبيرٍ من المُرُونة عندَ التَّأريخ للوحدات المُعجميَّة ومعانيها. فبالإضافة إلى أنَّ العربيَّةَ لم تبرَح صُورتَها المنطوقةَ لمئات السِّنين، فقد ضاعَ الكثيرُ من تُراثِها المكتوب، لاسيَّما التُّراثُ المُدَوَّنُ في سني التَّدوين الأولى من عُمرِها المديد.

وتجدرُ الإشارةُ إلى وجود علاقةٍ وطيدةٍ بينَ المعلومات التَّاريخيَّة والمعلومات التَّأثيليَّة الَّتي سَبَقَت الإشارةُ إليها. ولبيان ذلك يُمكنُ القَولُ إنَّهُ إذا كانت المعلوماتُ التَّأثيليَّةُ تُعنى بالأصل الَّذي كانَت عليه الكلمة، فالمعلومات التَّاريخيَّة تُعنى بتعقُّب التَّغيُّراتِ الحادثةِ على هذا الأصل عبرَ المراحل الزَّمنيَّة المُتعاقبة للُّغة؛ ومن ثمَّ، فإنَّها تخضعُ لقوانينِ التَّطوُّر اللُّغَويّ.

٢ , ٩- مُستويات الاستعمال (Usage Levels)

وهي المُستويات الَّتي تتحدَّدُ من خلالها حالةُ الوحدات المُعجميَّة وفقَ اعتباراتٍ مُجتَمَعيَّة مُتعدِّدة، منها على سبيل المثال: اعتِبار الشُّيُوع والإهمال (شائِع، مُطَّرِد، نادِر، شاذّ) واعتِبار الإباحة والحَظْر (محظُور، مُبتَذَل، سُوقِيّ، سَوادِيّ) واعتبار المكان (مَدَنِيّ، قَرَوِيّ، بَدَوِيّ) واعتبار الزَّمان (قديم، وسيط، حديث) ونحو ذلك. وبطبيعة الحال فإنَّ هذه المُستويات تخضعُ للعوامل الثَّقافيَّة والحضاريَّة والاجتماعيَّة للُّغة المُعَيَّنة.

٢ , ١٠- معلومات التَّكرار (Frequency Information)

هي المعلوماتُ الَّتي تُعنى بدرجة دوران الوحدات المُعجميَّة في اللُّغة. وتبدو الحاجةُ إليها عندَ الحاجة إلى التَّأكُّد من قَبُول المُجتَمَع اللُّغويّ لمُفرَداتٍ مُعَيَّنة أو رفضِه لها. ولهذه المعلومات أهمِّيَّةٌ في صناعة المعاجم التَّعليميَّة خُصُوصًا؛ حيثُ تُساعدُ مُتعلِّمي اللُّغة على الاهتداء بسُهولةٍ إلى أكثر المُفردات دورانًا في اللُّغة. والواقعُ أنَّ استخلاصَ هذه المعلومات لغاية الصِّناعة المُعجميَّة العربيَّة من الصُّعوبة بمكان، نظرًا لطبيعة اللُّغة العربيَّة الاشتقاقيَّة؛ حيثُ تتعدَّدُ أنماطُ الكلمات الدَّالَّة على وحدةٍ مُعجميَّةٍ مُعَيَّنة، على نحو ما نجدُ في الأنماط (كَتَبَ، يكتُب، تكتُب، تكتبون، ...) الدَّالَّة على الوحدة (كَتَبَ)، الأمر الَّذي يستدعي مُعالجاتٍ صرفيَّةٍ دقيقة.

٣- المُعالَجَة المُعجميَّة الآليَّة في مراحل الصِّناعة

تمرُّ الصِّناعة المُعجميَّة Lexicography بثلاث مراحلَ أساسيَّة، هي: الجمع، والتَّحرير، والنَّشر. وتتخلَّلُ كُلَّ مرحلةٍ مجموعةٌ من المراحل الفرعيَّة الَّتي تتحدَّدُ في ضوء طبيعة المُعجم والهدفِ منه. والواقعُ أنَّ المُعالَجَة الآليَّة لمُكوِّنات المُعجم اللُّغَويّ لا تنفردُ بها مرحلةٌ مُعَينة؛ لكنَّها تتمُّ في المراحل الثَّلاث، بأساليبَ وطرائقَ مُختلفة. وبيانُ ذلكَ على النَّحو الآتي:

٣,١- المُعالَجَة الآليَّة في مرحلة الجمع

تُعنى مرحلةُ الجمع ببناء المورد الأساسيّ الَّذي تتكوَّنُ عنهُ مادَّةُ المُعجم. وإذا أردنا أن نصنعَ مُعجمًا لُغويًّا مُمثِّلًا لواقع اللُّغة الطَّبيعيَّة، فمن البدهيّ أن نجمعَ مادَّتَهُ من المُجتمَع اللُّغويّ. ووفقًا لضوابط الصِّناعة المُعجميَّة الحديثة، فإنَّنا نصنعُ هيكلًا لمادَّة المُعجم [الَّتي تُمثِّلُ موردَهُ الرَّئيس] في صُورة مُدوَّنةً لُغويَّة Linguistic Corpus. ويُشارُ بالمُدوَّنة اللُّغويَّة إلى كُتلةٍ غير مُنتظمةٍ من النُّصُوص المُستمدَّة من واقع اللُّغة المُعَيَّنة، تُجمَعُ وفقَ أساليبَ إحصائيَّة، بما يتناسبُ مع الأهداف المنشودة، ويُمكنُ التَّحكُّمُ في بياناتها بالإضافة أو الحذف أو التَّعديل.

في الصِّناعة المُعجميَّة ينبغي أن تكونَ المُدوَّنةُ اللُّغويَّةُ تمثيلًا حقيقيًّا للُغة المُجتمَع؛ سواءٌ اللُّغة المُستخدَمة فعليًّا، أم الَّتي كانَت مُستخدمة في حِقبةٍ زمنيَّة أو مكانٍ مُعَيَّن. لهذا، فإنَّها تشتملُ على أعدادٍ هائلةٍ من الكلمات، تصلُ إلى الملايين، وتتجاوزُ الآحادَ إلى عشرات ومئات الملايين في بعض الأحيان. وبطبيعة الحال، فإنَّ مُعالجةَ هذه الأعداد من الكلمات ستكونُ بحاجةٍ إلى وقتٍ وجُهدٍ كبيرَين، ممَّا يستدعي توظيفُ الآلة على نحوٍ تتحقَّقُ معه الدِّقَّةُ والسُّرعة.

تستدعي المُعالَجَة الآليَّة لمُكوِّنات المُعجم في مرحلة الجمع إخضاع المُدوَّنة للمُعالَجَة عبرَ مُستويَين، هما: مُستوى مُعالَجَة المباني، ومُستوى مُعالَجَة المعاني؛ وذلك باستخدام أدواتٍ مُعيَّنة (سيأتي الحديثُ عنها لاحقًا). وينتجُ عن المُستوى الأوَّل: (المداخل والوحدات المُعجميَّة، ومعلومات التَّكرار، والمعاني الوظيفيَّة)، وعن المُستوى الآخَر: (المعاني المُعجميَّة، والحُقول والعلاقات الدِّلاليَّة، والمعلومات التَّاريخيَّة والتَّأثيليَّة، والشَّواهد، ومُستويات الاستعمال).

٣,٢- المُعالَجة الآليَّة في مرحلة التَّحرير

يقومُ تحريرُ المعاجم الحديثة على نظام التَّحرير المُعجَميّ Dictionary Writing System أو ما يُعرَفُ بنظام صناعة المُعجم Dictionary Production /Publishing System. وهو بيئةٌ حاسُوبيَّةٌ تفاعُليَّة، تشتملُ على قاعدة بياناتٍ للمعلومات المُعجميَّة وأداةٍ للتَّحرير المُعجميّ وأدواتٍ لإدارة الموارد اللُّغَويَّة والمُعجَمِيَّة والحاسُوبيَّة المُستخدمة في صناعة المُعجم، بالإضافة إلى الواجهة التَّفاعُليَّة لنظام التَّحرير. ويُساعدُ نظامُ التَّحرير المُعجَميّ على إخراج هيكل المُعجم في صيغةٍ قياسيَّةٍ ومُنتظمة تتوحَّدُ فيها مناهج المُحرِّرين المُشاركينَ في صناعة المُعجم.

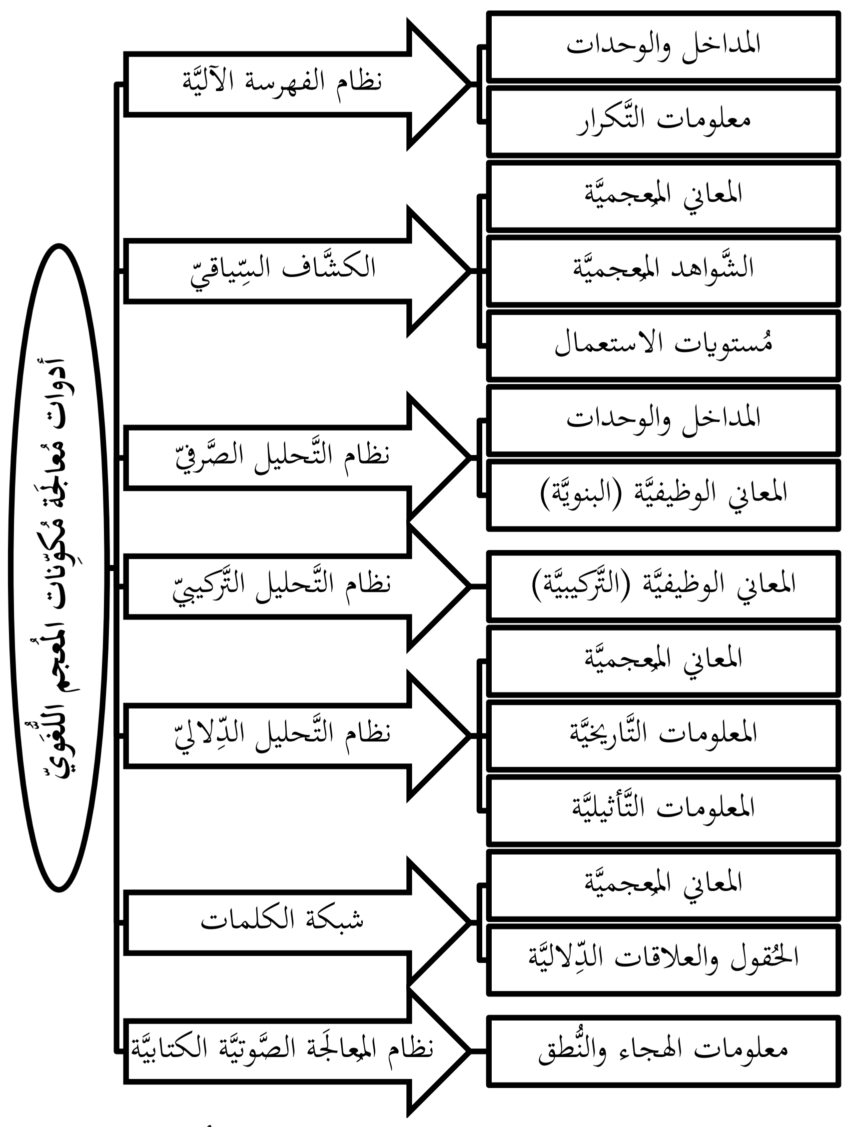
والواقعُ أنَّ أنظمةَ التَّحرير المُعجميّ ليسَت أدواتٍ للمُعالَجة الآليَّة المُعجميَّة بقَدر ما هي أنظمة للتَّحكُّم وإدارة البيانات. فهي تُساعدُ على استقبال مُخرَجات المُعالَجة المُعجميَّة ووضعها في هَيكلٍ مُنتظم، يُمكنُ التَّحكُّمُ فيه واستخلاص مُعطَياتِه وإحصاءاتِه. وتُساعدُ أنظمةُ التَّحرير كذلكً في ضبط الحُقول والعلاقات الدِّلاليَّة بينَ المُفرَدات، ونمذجة المعاني المُعجميَّة للمُفرَدات الَّتي تتبعُها. وتتفاوتُ درجةُ الإفادة من أنظمة التَّحرير المُعجميّ في إدارة المعلومات المُعجميَّة، وفقًا لطبيعة هذه المعلومات ودقَّة المُخرَجات النَّاتجة عن مُعالَجة المادَّة المُعجميَّة.

٣,٣- المُعالَجة الآليَّة في مرحلة النَّشر

إذا كانَ بالإمكان توظيفُ الآلة في استخلاص المعلومات المُعجميَّة في مرحلة الجمع، ثُمَّ توظيفها في التَّنظيم والتَّرتيب والهيكلة في مرحلة التَّحرير، فإنَّ دورَها يتجاوزُ الجوانبَ العلميَّة على الجوانب الفنِّيَّة في مرحلة النَّشر؛ حيثُ يكونُ التَّركيزُ على إخراج المادَّة المُعجميَّة في صورةٍ مُناسبةٍ للمُستخدمين، تتحقَّقُ معها أهدافُ الصِّناعة. لهذا فإنَّ المُعالجة الآليَّةَ لُمكوِّنات المُعجم في مرحلة النَّشر لا ترتبطُ بالمُكوِّنات الأساسيَّة بشكلٍ مُباشر؛ لكنَّها ترتبطُ بالمُكوِّنات الَّتي تتفرَّعُ عنها. وعلى سبيل المثال، تُوَجَّهُ المُعالَجَةُ الآليَّة إلى ضبط الوسائط المُتعدِّدة Multimedia الَّتي تتبعُ الوحدات المُعجميَّة ومعانيها، كالأشكال التَّوضيحيَّة والرُّسومات الَّتي تُستَخدَمُ في شرح المعنى [في المعاجم الورقيَّة] والصَّوتيَّات والمرئيَّات [في المعاجم المنطوقة]؛ كما يُستفادُ منها في ضبط مُخرَجات المعاجم المُوجَّهة إلى ذوي الاحتياجات الخاصَّة.

٤- من أدوات المُعالجَة الآليَّة لمكوِّنات المُعجم اللُّغَويّ

هُناكَ العديدُ من أدوات المُعالجَة الآليَّة الَّتي يُمكنُ الإفادةُ منها في مُعالجَة مُكوِّنات المُعجم اللُّغويّ. ويُوضِّحُ (الشَّكل ٦-١) أبرزَ أدوات المُعالجَة الآليَّة المُعجميَّة في اللُّغة العربيَّة، مع الإبانة عن المعلومات المُعجميَّة الَّتي يُمكنُ استخلاصُها عن عمل هذه الأدوات.



الشَّكل ٦-١: من أدوات المُعالجَة الآليَّة المُعجميَّة في اللُّغة العربيَّة

٤ , ١- نظام الفهرسة الآليَّة

يعملُ نظامُ الفهرسة الآليَّة Text Indexer على إعادة ترتيب المُفرَدات المُتضَمَّنة في المُدوَّنات اللُّغويَّة؛ كما يُساعدُ على إحصاء تردُّدات هذه المُفرَدات في النُّصُوص. ومن ثَمَّ، يرتكزُ دورُهُ في الصِّناعة المُعجميَّة على استخلاص المداخل والوحدات المُعجميَّة، وتعيين معلومات التَّكرار. وبطبيعة الحال، فإنَّ مُخرَجات هذا النِّظام تُوفِّرُ الوقتَ والجُهدَ، لا سيَّما عندَ التَّعامل مع أعدادٍ كبيرةٍ من النُّصُوص؛ لكنَّها في الوقت ذاتِهِ لا تُغني عن التَّدخُّل اليدويّ، لضبط مُخرَجات الآلة أوَّلًا، ثُمَّ التَّأكُّد من مُوافقة هذه المُخرَجات لطبيعة اللُّغة. وهناكَ ثلاثة أنواع من أنظمة الفهرسة الآليَّة الَّتي تدعمُ اللُّغةَ العربيَّة، وإن اختلَفَت في طرائق مُعالجتها للنُّصُوص، هي:

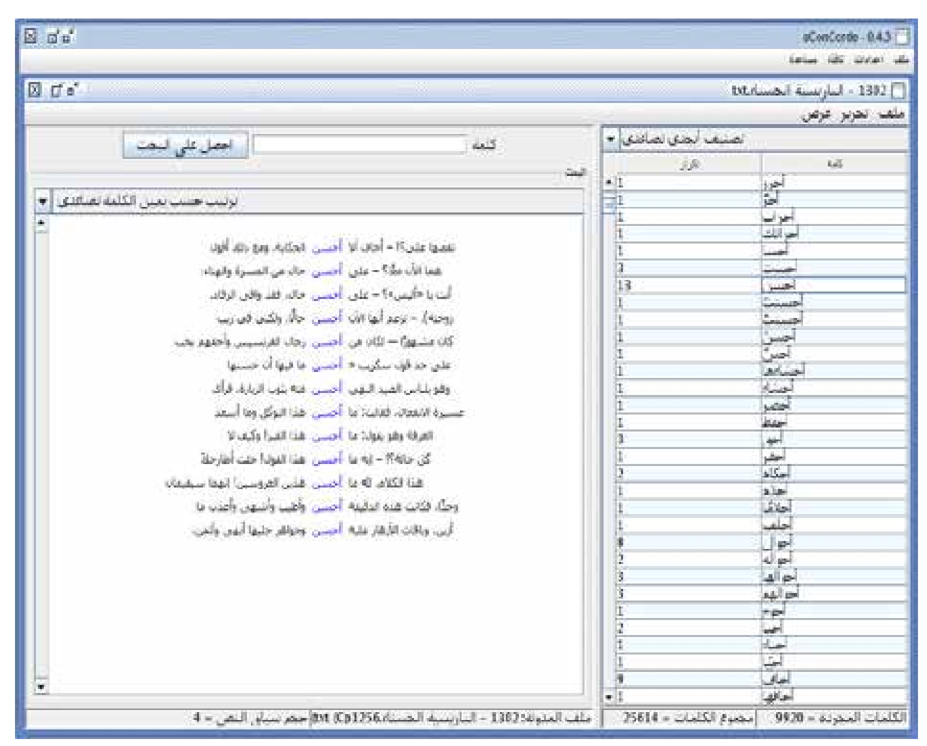
- **نظام الفهرسة الألفبائيَّة:** ويقومُ بترتيب المُفرَدات وفقًا للتَّرتيب الهجائيّ.
- **نظام الفهرسة الجذعيَّة:** ويقومُ بترتيب المُفرَدات وفقَ الجُذُوع Stems، بعدَ تجريدها من الزَّوائد [السَّوابق، واللَّواحق].
- **نظام الفهرسة الجِذريَّة:** ويقومُ بترتيب المُفرَدات وفقَ الجُذُور Roots، ويستدعي هذا النِّظامُ وُجودَ آليَّةٍ لإزالة الالتباس البنويّ بينَ المُفرَدات المُتَّفقة في شكلِها، المُختلفة في أصلِها؛ على نحو ما نجدُ مثلًا في كلمة [فَلَكَ]؛ حيثُ تدُلُّ في صورتها المُفرَدة على الفعل الماضي من الجذر (ف ل ك)، وتدلُّ أيضًا على حرف الحرف (ل) الَّذي تسبقُهُ فاءُ الاستئناف، وتلحقُ به كافُ الخطاب؛ ونحو ذلك.

والواقعُ أنَّ طبيعةَ اللُّغة تتحكَّمُ بصورةٍ كبيرةٍ في الجُهد البشريّ المبذول لاستخلاص المداخل والوحدات المُعجميَّة؛ بل إنَّ منهجَ التَّرتيب المُعجميّ في معاجم اللُّغة الواحدة يُحدِّدُ النِّظامَ الأكثرَ مُناسبةً للفهرسة. وعلى سبيل المثال، لن نكونَ بحاجةٍ إلى مجهودٍ كبيرٍ في استخلاص الوحدات المُعجميَّة في معاجم اللُّغات: الإنجليزيَّة، والألمانيَّة، والهولنديَّة، لأنَّها لُغات إلصاقيَّة. أمَّا العربيَّة، فتستدعي بذلَ جُهدٍ أكبرَ، نظرًا لطبيعتها الاشتقاقيَّة ونظامها الكتابيّ.

٤ , ٢- الكشَّاف السِّياقيّ

يُتمِّمُ الكشَّافُ السِّياقيّ Concordancer عملَ نظام الفهرسة الآليَّة؛ حيثُ يُعنى بتعيين سياقات كُلِّ مُفرَدةٍ على حِدة، بعدَ إخضاع جُملة نُصُوص المُدوَّنة للفهرسة الآليَّة. ويُساعدُ الكشَّافُ السِّياقيُّ بذلكَ على استخلاص المعاني المُعجميَّة من السِّياقات؛ إذ يعكسُ السِّياقُ معنى المُفردة حالَ استخدامِها في المُجتَمَع اللُّغويّ. ويُفيدُ الكشَّافُ السِّياقيُّ أيضًا في استخلاص الشَّواهد المُعجميَّة؛ كما يُساعدُ على تعيين مُستويات الاستعمال.

ونظرًا للتَّداخُل الواقع بينَ الكشَّاف السِّياقيّ ونظام الفهرسة الآليَّة، فكثيرًا ما نجدُهما في أداةٍ واحدةٍ للمُعالَجة، على النَّحو الوارد في (الشَّكل ٦-٢).



الشَّكل ٦-٢: نموذج من عمل المُفهرس الآليّ والكشَّاف السِّياقيّ (أداة aConCorde)

٤ ,٣- نظام التَّحليل الصَّرفيّ

المُحلِّل الصَّرفيُّ أحد اهمّ أدوات مُعالَجة المُفرَدات في اللُّغة العربيَّة؛ حيثُ يقومُ بتعيين جُذور الكلمات [أو أصولها] وتحليل كُلِّ كلمةٍ على عناصِرها الأوَّليَّة (الجُذوع، والسَّوابق، واللَّواحق)، بالإضافة إلى تعيين الفُرُوع Lemmata الَّتي تُعَدُّ شكلًا صرفيًّا للوحدات المُعجميَّة. وبهذا فإنَّ نظامَ التَّحليل الصَّرفيّ يُتمِّمُ أيضًا عملَ المُفهرس الآليّ؛ حيثُ يضعُ المداخل [المُمثَّلة في الجذور] والوحدات المُعجميَّة [المُمَثَّلة في الفُرُوع] في الصُّورة الَّتي ينبغي أن تظهرَ عليها. ومن ناحيةٍ أخرى، يُساعدُ المُحلِّلُ الصَّرفيُّ على تعيين المعاني الوظيفيَّة البنويَّة للوحدات المُعجميَّة. ويُوضِّحُ (الشَّكل ٦-٣) نموذجًا لمُخرَجات التَّحليل الصَّرفيّ في العربيَّة.

الدخل INPUT	الخرج OUTPUT							
	الكلمة المشكولة Voweled Word	السابق Prefix	الجذع Stem	نوع الكلمة Type	الوزن Pattern	الجذر Root	الحالة الإعرابية POS Tags	اللاحق Suffix
دوت	دَوَّتْ	#	دوت	فعل ماض مبني للمعلوم	فعَّت	دوي	ثلاثي مزيد مسند إلى الغائبة(هي) متعد ولازم	تْ: تاء التأنيث الساكنة
الغابة	الغَابَةُ	ال: التعريف	غابة	اسم جامد	فالة	غيب	مفرد مؤنث مرفوع في حالة التعريف	ة: تاء التأنيث
	الغَابَةَ	ال: التعريف	غابة	اسم جامد	فالة	غيب	مفرد مؤنث منصوب في حالة التعريف	ة: تاء التأنيث
	الغَابَةِ	ال: التعريف	غابة	اسم جامد	فالة	غيب	مفرد مؤنث مجرور في حالة التعريف	ة: تاء التأنيث
	الغَابَةُ	ال: التعريف	غابة	اسم فاعل	فاعة	غيب	مفرد مؤنث مرفوع في حالة التعريف	ة: تاء التأنيث
	الغَابَةَ	ال: التعريف	غابة	اسم فاعل	فاعة	غيب	مفرد مؤنث منصوب في حالة التعريف	ة: تاء التأنيث
	الغَابَةِ	ال: التعريف	غابة	اسم فاعل	فاعة	غيب	مفرد مؤنث مجرور في حالة التعريف	ة: تاء التأنيث
بأصوات	بِأَصْوَاتِ	ب: حرف الجر	أصوات	اسم جامد	أفعال	صوت	جمع مذكر مجرور في حالة الإضافة	#
	بِأَصْوَاتٍ	ب: حرف الجر	أصوات	اسم جامد	أفعال	صوت	جمع مذكر مجرور نكرة	#
السباع	السِّبَاعُ	ال: التعريف	سباع	اسم جامد	فِعَال	سبع	جمع مذكر مرفوع في حالة التعريف	#
	السِّبَاعَ	ال: التعريف	سباع	اسم جامد	فِعَال	سبع	جمع مذكر منصوب في حالة التعريف	#
	السِّبَاعِ	ال: التعريف	سباع	اسم جامد	فِعَال	سبع	جمع مذكر مجرور في حالة التعريف	#
	السِّبَاعُ	ال: التعريف	سباع	مصدر أصلي	فِعَال	سبع	مفرد مذكر مرفوع في حالة التعريف	#
	السِّبَاعَ	ال: التعريف	سباع	مصدر أصلي	فِعَال	سبع	مفرد مذكر منصوب في حالة التعريف	#
	السِّبَاعِ	ال: التعريف	سباع	مصدر أصلي	فِعَال	سبع	مفرد مذكر مجرور في حالة التعريف	#
الضارية	الضَّارِيَةُ	ال: التعريف	ضارية	اسم فاعل	فاعِيَة	ضرو	مفرد مؤنث مرفوع في حالة التعريف	ة: تاء التأنيث
	الضَّارِيَةَ	ال: التعريف	ضارية	اسم فاعل	فاعِيَة	ضرو	مفرد مؤنث منصوب في حالة التعريف	ة: تاء التأنيث
	الضَّارِيَةِ	ال: التعريف	ضارية	اسم فاعل	فاعِيَة	ضرو	مفرد مؤنث مجرور في حالة التعريف	ة: تاء التأنيث

الشَّكل ٦-٣: نموذج لمُخرَجات نظام التَّحليل الصَّرفيّ في العربيَّة (أداة Alkhalil)

٤ , ٤ – نظام التَّحليل التَّركيبيّ

يقومُ المُحلِّلُ التَّركيبيُّ Syntactic analyzer بتعيين أركان الجُملة الَّتي تُعرَفُ بأقسام الكلام Parts of Speech؛ كما يقومُ بتعيين وظائف هذه الأركان وتوصيفها من النَّاحية التَّركيبيَّة. والواقعُ أنَّ طبيعةَ التَّركيب في اللُّغة العربيَّة تستدعي تداخُلًا بينَ عمل المُحلِّلات التَّركيبيَّة وأدوات إزالة الالتباس الدِّلاليّ، نظرًا لعدَّة عواملَ، منها: قَبُول الجُملة العربيَّة الواحدة للتَّنوُّع بينَ الاسميَّة والفعليَّة [لا سيَّما في العربيَّة المُعاصرة]، وتغيُّر أطوال الجُملة بصورةٍ غَير مُطَّردة، وواقع الكتابة العربيَّة الَّتي تتجرَّدُ في كثيرٍ من الأحيان من علامات الضَّبط بالشَّكل، وغير ذلك. ويقتصرُ دَورُ المُحلِّل التَّركيبيّ في صناعة المُعجم على تعيين المعاني الوظيفيَّة التَّركيبيَّة لوحدات المُعجم.



الشَّكل ٦–٤: نموذج لآليَّة التَّحليل التَّركيبيّ في العربيَّة (Penn Arabic treebank)

٤ , ٥- نظام التَّحليل الدِّلاليّ

يرتكزُ عملُ المُحلِّل الدِّلاليّ Semantic Analyzer على المعاني المُعجميَّة للمُفرَدات في سياقاتِها. ومن النَّاحية العمليَّة في مُعالَجة مُكوِّنات المُعجم، ينبني نظامُ التَّحليل الدِّلاليّ على ما يُعرَفُ بآليَّة فَكِّ الالتباس الدِّلاليّ للكَلِماتِ Word Sense Disambiguation (WSD)). وَتَستَمِدُّ هذه الآليَّةُ فِكرَتَها الأَساسيَّةَ مِن المُتَصاحِبات [المُتلازمات] اللَّفظِيَّة Collocations الَّتي تَتَكَوَّنُ عن سِلسِلَةٍ مِن الكلمات، تَتَلازَمُ مُفرَداتُها في عَلاقَةٍ تَركِيبِيَّة، كَعَلاقَة الفِعل والفاعِل (مِثل: صاحَ الدِّيك، سادَ الصَّمْت)، وعَلاقَة الإضافة (مِثل: عابر سَبِيل، قاطِع طَريق). ويُستفادُ من نظام التَّحليل الدِّلاليّ في استخلاص المعاني المُعجميَّة، كمرحلةٍ تاليةٍ لمرحلة الكشف السِّياقيّ، خُصُوصًا عندَ التَّعامُل مع مجموعاتٍ كبيرةٍ من السِّياقات؛ كما يُستفادُ منهُ جُزئيًّا في استكشاف المعلومات التَّاريخيَّة والتَّأثيليَّة؛ حيثُ تدلُّ عليها سياقاتُ الوحدة المُعجميَّة. وعلى سبيل المثال، سترتبطُ الوَحدةُ (القِطار) باللُّغة العربيَّة القديمة إذا صاحَبَتها كلماتٌ مثل (الغَيث، العُشب، ...)، وبالعربيَّة الوسيطة إذا صاحَبَتها كلماتٌ مثل (القافلة، الإبل، ...)، وبالعربيَّة الحديثة والمُعاصِرة إذا صاحَبَتها كلماتٌ مثل (التَّذاكر، السِّكَّة، ...)، وهكذا.

م	السِّياق	الدِّلالة	الزَّمان	المُستوى
١	وَاستَبَحنا سَنامَ الأَرضِ إذ قَحِطَ القِطارُ	١	٢٢ ق.هـ	القديمة
٢	أنزلَ الدَّارَ المُبَنَّاةَ على سُقْيا القطار	١	٢٨٣ هـ	القديمة
٣	فلاَحَ لهم من بُعْدٍ قطارُ إبلٍ	٢	٣٤٦ هـ	الوسيطة
٤	ومن جملة ما كان معها قطار جمال مُحَمَّلَة	٢	٨٤٥ هـ	الوسيطة
٥	ثم سافرَ إلى بلده على **القطار الحديديّ**	٣	١٣٥٤ هـ	الحديثة
٦	هامَت تَمُرُّ على المرافِئ أو **محطَّات القطار**	٣	١٣٨٣ هـ	الحديثة

الجدول ٦-٤: نموذج للمُتصاحبات الدَّالَّة على مُفرَدات العربيَّة في آليَّة التَّحليل الدِّلاليّ

٤ , ٦- شبكة الكلمات

تُعَدُّ شبكةُ الكلمات WordNet موردًا رئيسًا في الصِّناعات اللُّغويَّة الحاسوبيَّة. وهي ليسَت أداةً للمُعالَجة الآليَّة بالمفهوم الدَّقيق؛ لكنَّها مَصدرٌ غَنيٌّ بالمعارف، مُعَدٌّ سَلفًا للمُساعدة في المُعالَجة الدِّلاليَّة والمُعجميَّة. ويُمكنُ تعريفُ شبكة الكلمات بأنَّها

قاعدة بيانات مُعجميَّة دلاليَّة Lexical-Semantic Database، تُصَنَّفُ المُفرَداتُ فيها إلى أقسام الكلام، ثُمَّ تُصَنَّفُ في سلاسلَ ومجمُوعاتٍ من المُتَرادفات Synonyms أو المفاهيم Synsets الَّتي تتَّفقُ في معانيها وتترابطُ فيما بينها بشبكةٍ من العلاقات الدِّلاليَّة. وبهذا التَّوصيف، يُمكنُ الإفادةُ من شبكات الكلمات عندَ مُعالجَة مُكوِّنات المُعجم اللُّغويّ في تتبُّع المعاني المُعجميَّة والاستدلال عليها؛ كما يُستفادُ من شبكات الكلمات على نطاقٍ واسعٍ في تعيين الحُقُول والعلاقات الدِّلاليَّة.

الشَّكل ٦-٥: نموذج لشبكة الكلمات في العربيَّة (Arabic WordNet)

٤, ٧- نظام المُعالجَة الصَّوتيَّة الكتابيَّة

الفونيم Phoneme هو وَحدة التَّحليل الصَّوتيّ في اللُّغات الطَّبيعيَّة في صُورتِها المنطوقة، ويُقابلُهُ في الصُّورة المكتوبة (الجرافيم Grapheme) الَّذي يُعبِّرُ عن أشكال المحارف. ويُستَخدَمُ المُحلِّل الصَّوتيّ الكتابيّ Phono-grapheme Analyzer في تعيين معلومات الهجاء والنُّطق، عَبرَ مُحاكاة القواعد القياسيَّة لأصوات اللُّغة ومحارِفها. وفي اللُّغة العربيَّة يبدو هذا النِّظامُ في صُورة ما يُعرَفُ بآليَّةٍ رَومنة الحُرُوف العربيَّة

Romanization of Arabic؛ أي: كتابتها باحرُفٍ رومانيَّة [أو لاتينيَّة]؛ حيثُ تحلُّ محارفُ الألفبائيَّة الصَّوتيَّة الدَّوليَّة مكانَ محارف اللُّغة العربيَّة؛ وفي بعض الأحيان، تكونُ هذه الآليَّة مُلحَقة بأنظمةٍ أخرى للتَّحليل الصَّرفيّ أو التَّركيبيّ.

م	الوحدات المُعجميَّة	قِسم الكلام	الصُّورة المنطوقة	الصُّورة المكتوبة
1	كتبَ	فِعل	*kAtAbA*	كَتَبَ
٢	عاهدَ	فِعل	*ʕÆhADA*	عاهَدَ
٣	استأثرَ	فِعل	*ʔiStAʔṯARA*	اسْتَأْثَرَ
٤	اللَّيل	اسم	*ʔAl-lAyl-ʊn*	لَيلٌ
٥	الَّذي	اسم	*ʔAllAḏi:*	الَّذي
٦	الوُدّ	اسم	*ʔAl-wʊDD-ʊn*	وُدٌّ
٧	النِّساء	اسم	*ʔAn-niSÆʔ-ʊ*	نِساءُ
٨	إنْ	حرف / أداة	*ʔin*	إنْ
٩	أنْ	حرف / أداة	*ʔAn*	أَنْ
١٠	كَي	حرف / أداة	*kAy*	كَيْ

الجدول ٦-٥: نماذج مُخرَجات المُعالجَة الفونوجرافيَّة [الصَّوتيَّة الكتابيَّة]

٥- أفكارٌ بحثيَّة لأطرُوحاتٍ علميَّةٍ ودراسات مُستقبليَّة

نَظَرًا لنُدرة الدِّراسات الَّتي كُتِبَت بالعَرَبِيَّة عن المُعالجَة المُعجميَّة الآليَّة وحداثة مناهج دراستها على اللُّغة العربيَّة، سنُحاوِلُ - فيما يأتي - أن نعرِضَ لبعضِ الأفكار البحثِيَّة الَّتي قد تَصلُحُ لإنجاز أطروُحاتٍ علميَّةٍ ودراساتٍ مُستقبليَّة للباحِثين في علوم اللُّغة والمُعجم.

٥,١- موضُوع الفِكرة الأولى:

أنساق المداخل والوَحدات في المُعجم العربيّ المُعاصِر

«دراسة لُغويَّة إحصائيَّة»

- **فكرة الدِّراسة:**

تقومُ فكرةُ الدِّراسة على حَصر وإحصاء أنماط المداخل والوحدات المُعجميَّة في بعض المعاجم العربيَّة المُعاصِرة أو أحدها؛ وتعيين الأنساق المُختلفة للمداخل حينَ تكونُ أصُولًا أو جُذورًا، وللوحدات حينَ تكونُ مُفرَدةً أو مُرَكَّبة؛ ثُمَّ ضبط هذه الأنساق على أساسٍ منهجيٍّ يُساعدُ في تقويم صناعة المُعجم العربيّ.

- **مادَّة الدِّراسة:**

مُعجم [أو أكثر] للُّغة العربيَّة المُعاصِرة.

- **الأسئِلة البَحثِيَّة:**

  - ما أنساقُ المداخل والوحدات في المُعجم العربيّ المُعاصِر؟
  - كيفَ يُمكنُ استخلاصُ منهجيَّة لضبط بنية هذه الأنساق في صناعة المُعجم؟

- **منهج الدِّراسة، ومجال البحث:**

تقومُ الدِّراسةُ المقتَرَحةُ على المنهج الوصفيّ التَّحليليّ، ويَتَنَوَّعُ مجالُ البحث بين الإحصاء اللُّغَويّ والصِّناعة المُعجميَّة.

٥,٢- موضُوع الفِكرة الثَّانية:

نحوَ مُعجمٍ حاسُوبيٍّ تفاعُليّ لذوي الاحتياجات الخاصَّة

«المنهج والنَّموذَج»

- **فكرة الدِّراسة:**

تقومُ فكرةُ الدِّراسة على وضع منهجيَّةٍ قابلةٍ للتَّطبيق لبناء وتطوير مُعجم لُغويٍّ مُعاصِرٍ، يعتمدُ على تقنيات الآلة في مراحلِ بنائِهِ الثَّلاثة (الجمع، والتَّحرير، والنَّشر)، ويظهرُ في بيئةٍ حاسُوبيَّةٍ تفاعُليَّةٍ، بهدف مُساعدة ذوي الاحتياجات

الخاصَّة من فاقدي القُدرة على السَّمع أو الإبصار على فهم اللُّغة وتعلُّم مُفرَداتها وإدراك نظامِها الـمُعجميّ، مع تقديم نموذجٍ حاسُوبيٍّ يُثبتُ الباحثُ من خلالِه فاعليَّةَ المنهج.

- **مادَّة الدِّراسة:**

مُعجمٌ عربيٌّ مُعاصِرٌ [أو مجموعة من المعاجم]، ومُدوَّنة لُغويَّة عامَّة للعربيَّة الـمُعاصِرة، وقاعدة بيانات دلاليَّة مُعجميَّة.

- **الأسئِلة البَحثِيَّة:**
  - ما ضوابطُ صناعة مُعجمٍ عربيٍّ لذوي الاحتياجات الخاصَّة؟
  - إلى أيِّ مدًى يُمكنُ بناءُ وتطوير مُعجمٍ عربيٍّ مُعاصِرٍ في بيئةٍ حاسُوبيَّةٍ تفاعُليَّة؟
  - كيفَ يُمكنُ توجيهُ الفئات المُستهدَفة من تعلُّم المُفرَدات ومعرفة المعلومات المُعجميَّة عَبرَ الآلة، دونَ وسيطٍ بشريّ؟

- **منهج الدِّراسة، ومجال البحث:**

تقومُ الدِّراسةُ المقترَحةُ على المنهج الوصفيّ، ويَتنَوَّعُ مجالُ البحث بين اللِّسانيَّات الحاسُوبيَّة والصِّناعة المُعجميَّة.

٥,٣- موضُوع الفِكرة الثَّالثة:

التَّوليد الآليّ للمعاني في معاجم اللُّغة العربيَّة المُعاصِرة

«مُعالجَة حاسُوبيَّة في ضوء تقنيات إزالة الالتباس الدِّلاليّ»

- **فكرة الدِّراسة:**

تُعالِجُ الدِّراسةُ الـمُقترَحة أحدَ أهمِّ إشكالات الصِّناعة الـمُعجميَّة الحديثة؛ ونعني إشكالَ استخلاص المعاني الـمُعجميَّة من السِّياقات. وتقومُ الفكرةُ على بناء مُدوَّنةٍ لُغويَّةٍ للعربيَّة الـمُعاصِرة، ثُمَّ إخضاع مادَّتها للتَّدريب والتَّجريب على تقنيات إزالة الالتباس الدِّلاليّ، والإبانة عن فاعليَّة هذه التِّقنيات في استخلاص المعاني وتوليدها آليًّا، وتقييم الـمُخرَجات استنادًا إلى معاجمَ عربيَّة مُنجَزةٍ سلفًا.

- **مادَّة الدِّراسة:**

مُدَوَّنةٌ لُغَوِيَّةٌ مكتُوبة (نَصِّيَّة) مُستَمَدَّةٌ من نُصُوص اللُّغة العربيَّة المعاصِرة، لتكونَ مادَّةً للتَّدريب والتَّجريب، ومُعجم عربي مُعاصِر، ليكونَ مادَّةً للتَّقييم.

- **الأسئِلة البَحثِيَّة:**

  - ما شكلُ المعنى في المعاجم العربيَّة الـمُعاصِرة، وما سِماتُه، وما هي القرائن اللُّغويَّة الَّتي يُمكنُ الاستدلالُ بها على المعاني الـمُعجميَّة للمُفرَدة الواحدة؟
  - كيفَ يُمكنُ توظيفُ تقنيات إزالة الالتباس الدِّلاليّ في استخلاص المعنى وتوليده؟
  - إلى أيِّ مدًى يُمكنُ توظيفُ هذه التِّقنيات في مُعالَجَة إشكالات صناعة الـمُعجم؟

- **منهج الدِّراسة، ومجال البحث:**

تقومُ الدِّراسةُ المقتَرَحةُ على المنهج الوصفيّ، ويَتَنَوَّعُ مجالُ البحث بين اللِّسانيَّات الحاسُوبيَّة ولسانِيَّات المدَوَّنة والصِّناعة المُعجميَّة.

٥, ٤- موضُوع الفِكرة الرَّابعة:

نحوَ قاعدة بيانات شَبَكِيَّة دلاليَّة لأغراض الصِّناعة المُعجميَّة العربيَّة

«دِراسةٌ لُغَوِيَّة حاسُوبيَّة»

- **فكرة الدِّراسة:**

تسعى الدِّراسةُ الـمُقتَرَحة إلى بناء وتطوير قاعدة بيانات من الحُقُول والعلاقات الدِّلاليَّة، لتكونَ موردًا للصِّناعة الـمُعجميَّة العربيَّة. وتقومُ الفكرةُ على إعداد ما يُشبهُ الهيكلَ أو النَّموذجَ الـمُصَغَّر لشبكة كلماتٍ WordNet تستمدُّ مادَّتَها ومُعطَياتِها من واقع اللُّغة العربيَّة الـمُعاصِرة، لتكونَ قادرةً على تمثيل المعاني الـمُعجميَّة على نحوٍ دقيق، مع تقديم نموذجٍ مُعجميٍّ مُستَخلَصٍ منها، للتَّأكُّد من مُناسبَتِها لأهداف الصِّناعة.

- **مادَّة الدِّراسة:**

شبكة الكلمات العربيَّة Arabic WordNet، وعدد من شبكات الكلمات والأنطولوجيات الـمُنجَزة للُّغات الأخرى.

- **الأسئِلة البَحثِيَّة:**
  - كيفَ نبني قاعدة بيانات دلاليَّة مُعجميَّة للعربيَّة؟
  - كيفَ يُمكنُ توظيفُ شبكة الكلمات في تطوير قاعدة البيانات؟
  - ما العلاقاتُ الأساسيَّة في قاعدة البيانات المنشودة، وكيفَ يُمكنُ توليدُها آليًّا، وإلى أيِّ مدًى يُمكنُ افادةُ من مُخرَجات التَّوليد الآليّ؟

- **منهج الدِّراسة، ومجال البحث:**

تقومُ الدِّراسةُ المقتَرَحةُ على المنهج الوصفيّ، ويَتَنَوَّعُ مجالُ البحث بين اللِّسانيَّات الحاسُوبيَّة وعلم الدِّلالة والصِّناعة الـمُعجميَّة.

٥,٥- موضُوع الفِكرة الخامسة:

بناء وتطوير نظام مفتوح المصدر للتَّحرير المُعجميّ العربيّ

«مُعالَجة حاسُوبيَّة»

- **فكرة الدِّراسة:**

مع وُجود عددٍ من أدوات التَّحرير الـمُعجميّ، إلَّا أنَّ اكثرَها لا يدعمُ اللُّغةَ العربيَّةَ دعمًا كاملًا، ولا يتناسبُ مع بنيتِها الـمُعجميَّة. لهذا، فإنَّ الدِّراسةَ الـمُقتَرَحة تسعى إلى بناء وتطوير نظامٍ للتَّحرير الـمُعجميّ العربيّ، يتناسبُ مع طبيعة اللُّغة العربيَّة من ناحية، ويكونُ مفتوحَ المصدر Open Source من ناحيةٍ أخرى. والهدفُ من ذلك تمكينُ الباحثينَ من استخدام نظام تحريرٍ مُعجميٍّ عربيٍّ بلا قُيُود، وتمكين العاملينَ في حوسَبة اللُّغة والعربيَّة وتقنياتِها من تطوير ذلكَ النِّظام وإثرائِه.

- **مادَّة الدِّراسة:**

مجموعة من المعاجم العربيَّة الـمُعاصِرة، وقاعدة بيانات مُعجميَّة قياسيَّة للتَّدريب.

- **الأسئِلة البَحثِيَّة:**
  - ما المواصفات القياسيَّة لنظام التَّحرير الـمُعجميّ العربيّ؟
  - إلى أيِّ مدًى يُمكنُ الإفادةُ من أنظمة التَّحرير الـمُعجميّ في صناعة مُعجمٍ عربيّ؟
  - ما الخُطوات المنهجيَّة لبناء وتطوير نظام مفتوح المصدر للتَّحرير الـمُعجميّ العربيّ، وما الرُّؤية الـمُستقبليَّة لدعم هذا النَّوع من أنظمة التَّحرير؟

- **منهج الدِّراسة، ومجال البحث:**

تقومُ الدِّراسةُ المقتَرَحةُ على المنهج الوصفيّ، ويَتَنَوَّعُ مجالُ البحث بين اللِّسانيَّات الحاسُوبيَّة والصِّناعة الـمُعجميَّة.

ببليوجرافيا مرجعيَّة

١. البعلبكِّي (رَوحي): المَورد العربيّ، دار العلم للملايين، بيروت، ط١، ٢٠١٦.

٢. السَّعيد (المُعتزّ بالله): حوسبة المعجم التَّاريخيّ للُّغة العربيَّة، مجلَّة اللِّسان العربيّ، مكتب تنسيق التَّعريب بالرِّباط، العدد ٧٤، ٢٠١٤.

٣. السَّعيد (المُعتزّ بالله): علم الدِّلالة ونظريَّة المعنى، دار الهاني، القاهرة، ٢٠١٢.

٤. السَّعيد (المُعتزّ بالله): نَمذَجة المُعجم العربيّ، مجلَّة اللِّسان العربيّ، مكتب تنسيق التَّعريب بالرِّباط، العدد ٧٦، ٢٠١٦.

٥. عُمَر (أحمد مُختار): صِناعة المُعجم الحديث، عالَم الكُتُب، القاهرة، ط٢، ٢٠٠٩.

٦. عُمَر (أحمد مُختار): في علم الدلالة، دار المعرفة الجامعية، الإسكندرية، ١٩٩٧.

٧. عُمَر (أحمد مُختار)، وآخرون: المُعجم العربيّ الأساسيّ، المُنظَّمة العربيَّة للتَّربية والثَّقافة والعُلُوم، لاروس، ١٩٨٩.

٨. مجمع اللُّغة العَرَبِيَّة بالقاهِرة: المُعجَم الوسيط، مكتبَة الشّروق الدَّولِيَّة، القاهِرَة، ط٣، ٢٠٠٤.

9. Abel, A. (2012). Dictionary writing systems and beyond. In: Granger, S. & Paquot, M. (ED). Electronic Lexicography. Oxford University Press, 2012.

10. Black, W., Elkateb, S., Rodriguez, H, Alkhalifa, M., Vossen, P., Pease, A. and Fellbaum, C., (2006). Introducing the Arabic WordNet Project, in Proceedings of the Third International WordNet Conference, GWC 2006, South Jeju Island, Korea, January 22-26, 2006.

11. Cantos-Gomez, P., Almela-Sanchez, M. (2019). Lexical Collocation Analysis: Advances and Applications. Springer Nature Customer Service Center Gmbh.

12. Čibej, J., Gorjanc, V., Kosem, I., Krek, S. Lexicography in Global Contexts: Book of Abstracts. Ljubljana University Press, Faculty of Arts.

13. Dash, N. S., Ramamoorthy, L. (2018). Utility and Application of Language Corpora. Springer.

14. Draper, C. (2018). Synonyms and Antonyms. Achieve2day.
15. Fuertes-Olivera, P. A. (2017). The Routledge Handbook of Lexicography. Routledge.
16. Hagberg, G. (2018). Meaning and Interpretation: Wittgenstein, Henry James, and Literary Knowledge. Cornell University Press.
17. Harm, V., Lobenstein-Reichmann, A., Diehl, G. (2019). Wortwelten: Lexikographie, Historische Semantik und Kulturwissenschaft. Walter de Gruyter GmbH & Co KG.
18. Heusinger, K., Maienborn, C., Portner, P. (2019). Semantics - Typology, Diachrony and Processing. Walter de Gruyter GmbH & Co KG.
19. Mel'čuk, I. (2015). Semantics: From meaning to text. John Benjamins Publishing Company.
20. Nielsen, S. (2008), "The Effect of Lexicographical Information Costs on Dictionary Making and Use", Lexikos, 18: 170–189.
21. Opland, J. (2019). Lexicography: Notes on Xhosa Lore and Language (1909-1934) - Robert Godfrey. University of Kwazulu-Natal Press.
22. Pustejovsky, J., Batiukova, O. (2019). The Lexicon. Cambridge University Press.
23. Reqqass, M., Lakhouaja, A., Mazroui, A., Bebah, M. (2014). "Conception et réalisation d'un système de production de dictionnaires arabes respectant la norme LMF", 5th International Conference on Arabic Language Processing. CITALA 2014. November 26-27. 2014, Oujda, Morocco.
24. Šipka, D. (2019). Lexical Layers of Identity: Words, Meaning, and Culture in the Slavic Languages. Cambridge University Press.
25. Zhekov, M. A. (2018). Lexicography in Action: Designing an English-Romanian Glossary of English Criminal Law for Interpreters and Legal Professionals Using Electronic Corpora. CreateSpace Independent Publishing Platform.

الباحثون

الدكتور/ مُحسِن عبد الرَّازق علي رشوان



يشغل منصب أستاذٍ بقسم الإلكترونيات والاتصالات الكهربائية في كُلِّيَّة الهندسة - جامعة القاهرة. تخرَّجَ عام ١٩٧٧ وكان الأول على دفعته، وحصلَ على ثلاثة ماجستيرات، ثم على الدُّكتوراه من جامعة كوين بكندا؛ أشرف على أكثر من مائة رسالة ماجستير ودكتوراه. يدير الشَّركة الهندسيَّة لتطوير النُّظُم الرَّقمِيَّة RDI المتخَصِّصة في مجال تقنيات اللُّغة العربيَّة.

الدكتور/ المُعتزّ بالله السَّعيد طه



أستاذ الدِّراسات اللُّغويَّة المُساعد بجامعة القاهرة، وأستاذ اللِّسانيَّات الحاسوبيَّة المُشارك بمعهد الدَّوحة للدِّراسات العُليا، ومُنسِّق وَحدة الموارد المُعجميَّة بمشروع مُعجم الدَّوحة. نَشَرَ نحوَ ثلاثينَ ورقة علميَّةٍ، بالإضافة إلى عددٍ من الكتب في المُعجَمِيَّة العَرَبيَّة والدِّراسات اللُّغَوِيَّة المُعاصِرة، وأسهمَ في أكثر من عشرة مشرُوعاتٍ بحثيَّةٍ دوليَّةٍ في ميادين مُعالَجة اللُّغات الطَّبيعيَّة. حصلَ على عددٍ من الجوائز في مَيدان تخصُّصه، منها: جائزة (ألكسو ALECSO) للإبداع والابتكار في «المَعلُوماتيَّة والمُعالَجة الآليَّة للُّغة العربيَّة»، وجائزة راشد بن حُمَيد للعُلُوم والثَّقافة.

الدكتور/ أحمد راغب أحمد



أُسْتَاذ الدِّراسات اللُّغَويَّة المُشَارِك، ورَئِيس قِسْم اللغة العربية وآدابها بالجامعة الإسلامية العالمية بماليزيا. حصلَ على درجة الدُّكتوراه من جامعة القاهِرة عام ٢٠٠٩. وَنَشَرَ عددًا من الأوراق البحثِيَّة في دوريَّاتٍ علميَّة؛ كما شارَكَ في العديد من المُؤتمرات الدَّوليَّة المَعنِيَّة بحوسبة اللُّغة؛ وله عددٌ من المُؤَلَّفات العِلميَّة.

الدكتور/ مُحمَّد عَطِيَّة مُحمَّد العَرَبيّ



حصل من جامعة القاهرة على بكالوريوس هندسة الاتصالات الكهربائية والإلكترونيات عام ١٩٩٥م، وعلى الماجستير في هندسة الحاسبات عام ٢٠٠٠م، ثم على درجة الدكتوراه في هندسة الاتصالات الكهربائية والإلكترونيات عام ٢٠٠٥م. عَمِلَ بالشَّركة الهندسيَّة لتطوير النُّظُم الرَّقمِيَّة RDI منذ يوليو ١٩٩٥م إلى نهاية ٢٠١٠م، وبينَ عامَيْ ٢٠٠٧م و ٢٠١٠م أستاذًا زائرًا في كلية الحاسبات وتقنيات المعلومات بالأكاديمية البحرية للعلوم والتكنولوجيا والنقل البحري - فرع القاهرة، ومخطِّطًا للبرمجيات بشركة «لوكسور تكنولوجي» الكندية (أُوكْفِيلْ/ أُونْتارْيُو) منذ ٢٠٠٩م إلى الآن، وخبيرًا للغويات الحاسوبية والبرمجيات لمشروع «معجم الدوحة التاريخي» بين عامَيْ ٢٠١٤م و ٢٠١٦م.

الدكتور/ سامح سَعد أبو المَجد الأنصاريّ



يَعمل أستاذًا للِّسانيَّات الحاسُوبِيَّة ورئيسًا لقسم الصَّوتِيَّات واللِّسانيَّات بكُلِّيَّة الآداب بجامعة الإسكندريَّة، ومديرًا لمركز اللُّغويَّات الحاسوبيَّة العربيَّة بمكتبة الإسكَندَرِيَّة. شارك في العديد من المشروعات العلمية ونَشَرَ العَدِيدَ من الأوراق البَحثِيَّة المَعنيَّة بحوسبة اللُّغة العَرَبِيَّة؛ وهو عُضوٌ بجمعيَّة اللِّسانيَّات العَرَبِيَّة بالولايات المُتَّحدة الأمريكِيَّة، وعضوٌ بمُؤَسَّسة لُغة الشَّبكات الدلالية الحاسوبية العالَمِيَّة بجنيف.

مباحث لغوية 00

مُقَدِّمةٌ في
حَوْسَبَة اللُّغَـــة العَرَبِيَّة

يُصدِر مركز الملك عبدالله بن عبدالعزيز الدولي لخدمة اللغة العربية هذا الكتاب ضمن سلسلة (مباحث لغوية)، وذلك وفق خطة عمل مقسمة إلى مراحل، لموضوعات علمية رأى المركز حاجة المكتبة اللغوية العربية إليها، أو إلى بدء النشاط البحثي فيها، واجتهد في استكتاب نخبة من المحررين والمؤلفين للنهوض بعنوانات هذه السلسلة على أكمل وجه.

ويهدف المركز من وراء ذلك إلى تنشيط العمل في المجالات التي تُنَبِّه إليها هذه السلسلة، سواء أكان العمل علميا بحثيا، أم عمليا تنفيذيا، ويدعو المركز الباحثين كافة من أنحاء العالم إلى المساهمة في هذه السلسلة.

وتودّ الأمانة العامة أن تشيد بجهد السادة المؤلفين، وجُهد مُحرِّرَي الكتاب، على ما تفضلوا به من رؤى وأفكار لخدمة العربية في هذا السياق البحثي.

والشكر والتقدير الوافر لمعالي وزير التعليم المشرف العام على المركز، الذي يحث على كل ما من شأنه تثبيت الهوية اللغوية العربية، وتمتينها، وفق رؤية استشرافية محققة لتوجيهات قيادتنا الحكيمة.

والدعوة موجّهة إلى جميع المختصين والمهتمين للتواصل مع المركز؛ لبناء المشروعات العلمية، وتكثيف الجهود، والتكامل نحو تمكين لغتنا العربية، وتحقيق وجودها السامي في مجالات الحياة.

الأمين العام للمركز

أ. د. محمود إسماعيل صالح







